Université de Montréal

New simulation schemes for the Heston model

par

Jean-François Bégin

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)

en Mathématiques

Orientation Mathématiques appliquées

juin 2012

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

New simulation schemes for the Heston model

présenté par

Jean-François Bégin

a été évalué par un jury composé des personnes suivantes :

Louis G. Doray

(président-rapporteur)

Mylène Bédard

(directeur de recherche)

Patrice Gaillardetz

(co-directeur)

Jean-François Angers

(membre du jury)

Mémoire accepté le:

14 juin 2012

SOMMAIRE

Les titres financiers sont souvent modélisés par des équations différentielles stochastiques (ÉDS). Ces équations peuvent décrire le comportement de l'actif, et aussi parfois certains paramètres du modèle. Par exemple, le modèle de Heston (1993), qui s'inscrit dans la catégorie des modèles à volatilité stochastique, décrit le comportement de l'actif et de la variance de ce dernier.

Le modèle de Heston (1993) est très intéressant puisqu'il admet des formules semi-analytiques pour certains produits dérivés, ainsi qu'un certain réalisme. Cependant, la plupart des algorithmes de simulation pour ce modèle font face à quelques problèmes lorsque la condition de Feller (1951) n'est pas respectée.

Dans ce mémoire, nous introduisons trois nouveaux algorithmes de simulation pour le modèle de Heston (1993). Ces nouveaux algorithmes visent à accélérer le célèbre algorithme de Broadie et Kaya (2006) ; pour ce faire, nous utiliserons, entre autres, des méthodes de Monte Carlo par chaînes de Markov (MCMC) et des approximations.

Dans le premier algorithme, nous modifions la seconde étape de la méthode de Broadie et Kaya afin de l'accélérer. Alors, au lieu d'utiliser la méthode de Newton du second ordre et l'approche d'inversion, nous utilisons l'algorithme de Metropolis-Hastings (voir Hastings (1970)).

Le second algorithme est une amélioration du premier. Au lieu d'utiliser la vraie densité de la variance intégrée, nous utilisons l'approximation de Smith (2007). Cette amélioration diminue la dimension de l'équation caractéristique et accélère l'algorithme.

Notre dernier algorithme n'est pas basé sur une méthode MCMC. Cependant, nous essayons toujours d'accélérer la seconde étape de la méthode de Broadie

et Kaya (2006). Afin de réussir ceci, nous utilisons une variable aléatoire gamma dont les moments sont appariés à la vraie variable aléatoire de la variance intégrée par rapport au temps. Selon Stewart *et al.* (2007), il est possible d'approximer une convolution de variables aléatoires gamma (qui ressemble beaucoup à la représentation donnée par Glasserman et Kim (2008) si le pas de temps est petit) par une simple variable aléatoire gamma.

MOTS CLÉS : Volatilité stochastique, algorithmes de simulation, tarification de produits dérivés, MCMC, algorithme de Metropolis-Hastings, modèle de Heston, options asiatiques, approximation, loi gamma.

ABSTRACT

Financial stocks are often modeled by stochastic differential equations (SDEs). These equations could describe the behavior of the underlying asset as well as some of the model's parameters. For example, the Heston (1993) model, which is a stochastic volatility model, describes the behavior of the stock and the variance of the latter.

The Heston (1993) model is very interesting since it has semi-closed formulas for some derivatives, and it is quite realistic. However, many simulation schemes for this model have problems when the Feller (1951) condition is violated.

In this thesis, we introduce new simulation schemes to simulate price paths using the Heston (1993) model. These new algorithms are based on Broadie and Kaya's (2006) method. In order to increase the speed of the exact scheme of Broadie and Kaya, we use, among other things, Markov chains Monte Carlo (MCMC) algorithms and some well-chosen approximations.

In our first algorithm, we modify the second step of the Broadie and Kaya's method in order to get faster schemes. Instead of using the second-order Newton method coupled with the inversion approach, we use a Metropolis-Hastings algorithm.

The second algorithm is a small improvement of our latter scheme. Instead of using the real integrated variance over time p.d.f., we use Smith's (2007) approximation. This helps us decrease the dimension of our problem (from three to two).

Our last algorithm is not based on MCMC methods. However, we still try to speed up the second step of Broadie and Kaya (2006). In order to achieve this, we use a moment-matched gamma random variable. According to Stewart *et al.*

(2007), it is possible to approximate a complex gamma convolution (somewhat near the representation given by Glasserman and Kim (2011) when $T-t$ is close to zero) by a gamma distribution.

KEYWORDS: Stochastic volatility, Simulation schemes, Asset pricing, MCMC, Metropolis-Hastings algorithm, Heston model, Asian options, approximations, gamma distribution.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BSM	Black-Scholes-Merton
BK-I	Broadie and Kaya drift interpolation
c.d.f.	Cumulative distribution function
GA	Gamma approximation
GE	Gamma expansion
IG	Inverse gaussian
ODE	Ordinary differential equation
LHS	Left-hand side
MCMC	Markov Chain Monte Carlo
p.d.f.	Probability density function
PDE	Partial differential equation
QE	Quadratic exponential
RHS	Right-hand side
SDE	Stochastic differential equation
TG	Truncated gaussian
TV	Transformed volatility

NOTATION

$\stackrel{d}{=}$	Equal in distribution
$\|\alpha\|$	Absolute value of α
$\mathbb{C}$	Set of all complex numbers
$C(S(t),K,T-t,\Theta)$	A European call option price on S at time t with a maturity T and a strike price of K. The model's parameters (whatever the model is) are in the vector Θ
$\mathbb{E}$	Expected value
$\mathbb{I}_A(\cdot)$	Indicator function which equals 1 if the argument is in A and 0 if the argument is in A^C
$\mathscr{I}_\nu(z)$	A Bessel of the first kind random variable with ν degrees of freedom evaluated at z.
i	Imaginary number, $\mathrm{i}=\sqrt{-1}$
$\mathscr{N}(\mu,\sigma^2)$	A normal random variable of mean μ and variance σ^2
$\Phi(\cdot)$	c.d.f. of a normal distribution $\mathscr{N}(0,1)$
$\phi(\cdot)$	p.d.f. of a normal distribution $\mathscr{N}(0,1)$
$\mathbb{P}$	Real probability measure
$\mathbb{Q}$	Risk-neutral probability measure (also known as martingale measure)
$\mathbf{Re}$	Real part
$\mathbb{R}$	Set of all real numbers
$\mathbb{R}^+$	Set of all non-negative real numbers
$S^{(0)}(t)$	The price of a risk-free asset $S^{(0)}$ at time t
$\mathscr{U}(a,b)$	A uniform random variable of lower bound a and upper bound b

ACKNOWLEDGEMENTS

First, I would like to thank my two supervisors, Dr. Mylène Bédard and Dr. Patrice Gaillardetz for the help they provided me with during my research and the excellent advices they gave me. I enjoyed the discussions we had during the last years; they were stimulating and they helped me evolve, as a scholar and as a human being.

I would also like to thank the whole team at Université de Montréal: especially, the professors who gave me the knowledge and the strength I needed to write these words. You know who you are.

Since my youth, my parents have always been present for me. They helped me in many situations. Thank you for everything, Mom and Dad. Without both of you, I would not be here today.

Finally, as the saying goes: "behind every successful man, there is a woman". I take this opportunity to thank my life partner, Sabrina, for everything she has done for me during my Master's degree. Doing research is not always an easy task. However, you were always there for me, even during the darkest days. Thank you Sabrina for everything you have done.

INTRODUCTION

Financial stocks are often modeled by stochastic differential equations (SDEs). These equations describe the behavior of the underlying asset and some of the model's parameters. Nowadays, the models retaining our attention have stochastic volatility. They are known to allow for a better calibration to market data; they also efficiently capture the smile of the volatility observed in financial securities.

Until recently, the Black-Scholes-Merton (BSM) model (Black and Scholes (1973) and Merton (1973)) was widely used. Semi-closed formulas exist for this model, which make it very attractive from a practical point of view. However, the BSM model includes very coarse assumptions such as a constant volatility and a deterministic constant interest rate. These shortfalls, combined to several financial crashes and the introduction of new complex products, have forced the financial analysts to develop new models. Heston (1993) proposed a model based on the square root process with mean reversion to express variance (not volatility). This model became very popular and is widely used in practice since Heston managed to derive a semi-closed formula for the price of a European call option. Moreover, the variance process (square root process) is widely applied in finance since many analytic results are known about this SDE. For a concrete example, we refer the reader to Cox *et al.* (1985) short rate model.

Although, there is a formula for pricing the European call option under the Heston model. There is none for more complex products that require path dependency. Therefore, we must, in practice, use Monte Carlo simulation techniques to price theses products. Even if the Heston model is nearly twenty years old, efficient simulation procedures have interested only a handful of individuals. The

Euler-Maruyama and the Milstein (1978) schemes are the easiest to implement and the most efficient. These methods can be used with almost every SDE. However, with the Heston model, these techniques do not work very well when the time steps are long and when the Feller condition is not satisfied. Several people have tried to improve these sampling methods for the Heston model.

Lord *et al.* (2008) considered a large number of fixes for the Euler-Maruyama method. They finally proposed a new scheme called "full truncation" which seems to work surprisingly well when compared to the other fixes, according to empirical results. This scheme is built so that it minimizes the bias on European call option prices.

The Kahl and Jäckel (2006) method also falls in the same batch of improvements. They suggest discretizing the variance process V using an implicit Milstein scheme coupled with their proposed discretization method for the asset process. This integration scheme is based on four key elements: the interpolation of the drift of the price, the interpolation of the diffusion term of the price, the decorrelation of the diffusion term, and the inclusion of higher order Milstein terms. Unfortunatetly, according to Andersen (2007), this scheme has an important bias when the Feller condition is not satisfied.

On a different note, Broadie and Kaya (2006) proposed a method, said to be exact, to simulate the Heston model. Essentially, by generating a value for the variance, it is then possible to generate the conditional integrated variance over time given the bounds of the integral, using Fourier inversion techniques. In their paper, they derive the characteristic function of this integral in a semi-closed form. Thus, this characteristic function makes the simulation of this integral possible. Subsequently, one can easily recover the price. This method, though elegant, is not applicable in practice. The computational intensity required by this algorithm makes it almost unusable (see Haastrecht and Pelsser (2010)). Nonetheless, several researchers used the idea of Broadie and Kaya (2006) to develop their own techniques.

For example, Smith (2007) proposed an almost exact algorithm. He replaced the arithmetic and the geometric averages in the characteristic function by a

weighted average of these two means. Thus, the characteristic function depends on two variables instead of three. We can then cache the values of the characteristic function for each value of the two variables. This should theoretically speed up the simulation algorithm. However, this hotfix barely accelerates the Broadie and Kaya (2006) method: the Smith (2007) technique is about 70 times slower than the best methods according to Haastrecht and Pelsser (2010).

Still using the basics of Broadie and Kaya (2006), Glasserman and Kim (2011) focus on the simulation of the integrated variance over time given the bounds of the integral, they found that the integral of the variance can be explicitly represented by a sum of infinite mixtures of gamma random variables. Thus, they can generate this random variable more easily by truncating the infinite series. Consequently, the longest step of the Broadie and Kaya (2006) method becomes much faster.

Andersen (2007) takes a similar avenue. He presented two state-of-the-art methods based on moments matching and efficient approximations of the variance. The idea behind his first scheme, called truncated Gaussian (TG hereafter), is to match the moments of a Gaussian density where the probability under zero is assigned to a Dirac delta function at the origin. Thus, the pattern will reproduce the asymptotic behavior of the variance process for large values. For smaller values of the variance process, V, the scheme will approximate the chi-squared distribution. In his second scheme, called quadratic exponential (QE hereafter), Andersen (2007) fixes a small problem of his TG method: when V is small, the Gaussian density decreases too quickly. To overcome this problem, he divides the problem in two parts (small or large V). Thus, he can choose a new density for small values of V based on an exponential and a Dirac delta function. This algorithm is considered by many, including Tse and Wan (2010), one of the best algorithms for simulating the Heston model. Though, according to Zhu (2008), it has a well-known problem: leaking correlation.

Zhu's (2008) method differs somewhat from the other methods in the literature. Instead of working with the variance SDE, he works with the root of the variance SDE, or what we call volatility. Using some approximations for the new

process' parameters and the Euler-Maruyama discretization, he shows that it is possible to obtain a method that gives very similar results to Andersen (2007). Moreover, his technique does not have the leaking correlation problem known in the QE method. However, this scheme does not work well when the Feller condition is violated.

Finally, Tse and Wan (2010) consider a biased approximation based on the inverse Gaussian distribution. They show that the integrated variance of the Heston model converges to an inverse Gaussian when the time step goes to infinity. Using this result, Tse and Wan (2010) develop an algorithm that uses an efficient approximation of the inverse Gaussian to simulate the Heston model.

The schemes we propose use the basics of Broadie and Kaya (2006); in order to speed up the latter scheme, we use MCMC algorithms to simulate the integrated variance for the first two algorithms. By caching the probability density function (p.d.f.) of the integrated variance over time and by using the famous Metropolis-Hastings algorithm, we are able to quickly sample the integrated variance. The main difference between the two proposed schemes is that in the first one, we use the Smith's approximation of the real p.d.f. instead of the real one derived by Broadie and Kaya.

The third one is based on a moment-matched gamma random variable. Glasserman and Kim (2011) show that the integrated variance over time can be represented by a sum of infinite mixtures of gamma random variables. Moreover, according to Stewart *et al.* (2007), it is possible to approximate a complex gamma convolution (somewhat near the representation given by Glasserman and Kim (2011)) by a simple gamma random variable where the moments are matched.

The contribution of this thesis is two-fold. First, we propose our new simulation methods for the Heston model. Then, we compare our new methods to some important schemes used in practice; these numerical tests will use realistic and challenging parameters (similar to those observed in market data). Our applications focus on implied model parameters that can be considered quite extreme.

This thesis is organized as follows: in Chapter 1, we first outline some preliminary materials which will be useful later. Then, in Chapter 2, we present in

detail the Heston dynamics and we introduce some important results about the processes involved. We introduce the MCMC basics in Chapter 3. Chapter 4 is devoted to an analysis of the most popular simulation schemes for the Heston model. In Chapter 5, we introduce our first MCMC-based algorithm and we compare it with those discussed earlier. In Chapter 6, we explain our second scheme and, again, compare it to the schemes available in the literature. Finally, Chapter 7 is dedicated to a third algorithm based on a gamma approximation. We also compare this third scheme to the ones available in the literature.

Chapter 1

PRELIMINARY MATERIAL

We discuss some important results from the two main fields considered in this thesis: finance and mathematics. The basics introduced in this chapter should help clarify the notions to be introduced subsequently.

1.1. Financial notions

We first describe the concept of arbitrage (which will guide every step of our methods). Then, we quote the Fundamental Theorem of Arbitrage Pricing. Finally, we define the concept of option. Most of these definitions come from Gauthier (2010b) and Gauthier (2010c).

Definition 1.1.1. *Arbitrage is the practice of taking advantage of a price difference between two or more markets. There are three conditions; if one of these is met, arbitrage is possible:*

(1) The same asset does not trade at the same price on two (or more) markets. This condition is also called "the law of one price".

(2) Two assets with identical cash flows do not trade at the same price.

(3) An asset with a known price in the future does not trade today at its future price discounted at the risk-free interest rate.

The existence of an arbitragre-free measure is important in asset pricing theory. Without this measure, called martingale measure or risk-neutral measure, different prices for the same product could be observed, thus creating arbitrage opportunities.

Theorem 1.1.1 (The Fundamental Theorem of Arbitrage Pricing)**.** *A financial market with time horizon T and price processes of the risky asset and riskless money market account given by $S(t)$, $t \in [0,T]$ and $S^{(0)}(t)$, $t \in [0,T]$, respectively, is arbitrage-free under the probability $\mathbb{P}$ if and only if there exists another probability measure $\mathbb{Q}$ such that*

(1) For any event A, $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$. We say this in the case that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent probability measures.

(2) The discounted price process, X, such that $X(t) = \frac{S(t)}{S^{(0)}(t)}$, $t \in [0,T]$ is a martingale under $\mathbb{Q}$.

A measure $\mathbb{Q}$ that satisfies 1 and 2 is known as a risk-neutral measure.

Proof. We refer the reader to Harrison and Pliska (1981) for a formal proof of Theorem 1.1.1. □

The price of an asset depends crucially on the risk of this particular asset, as investors typically demand more profit for bearing more uncertainty. Hence, in order to efficiently price a product, one would need to evaluate the physical probabilities (of an investor's beliefs) and find the discount factor associated with these probabilities.

It turns out that in a complete market (one in which the complete set of possible gambles on future states-of-the-world can be constructed with existing assets without friction) with no arbitrage opportunity there is an alternative way to do this calculation. One can compute, once and for all, the probabilities (i.e. risk-neutral) incorporating all investors risk premia, instead of adjusting for an investor's risk appetite. Once these risk-neutral probabilities are obtained, every asset can be priced by taking its expected payoff. Unlike risk-neutral probabilities, the actual real-world probabilities lead to a different adjustement (as they differ in riskiness) for every asset.

We now define some important financial products. These products shall be used later to assert our methods. These definitions are inspired from Hull (1989).

Definition 1.1.2. *A call option gives the holder the right to buy the underlying asset by a certain date for a certain price. A put option gives the holder the right*

to sell the underlying asset by a certain date for a certain price. The date specified in the contract is known as the maturity date and the price specified is known as the strike price.

Thus, the call option holder receives $\max(S(T)-K,0)$, where the price of the asset at maturity is $S(T)$ and K is the strike price. The put option holder receives $\max(K-S(T),0)$ at maturity. There exist many kinds of options; we now define some of them and explain the differences between them. European options can only be exerciced at the maturity date, while, American options can be exerciced at any time up to the maturity date. Most of the options traded are American, but they are obviously more expensive than the European options (or of equal value). There is a third popular kind of options called Asian options. The payoff of these options depends on the average price of the underlying asset during at least some part of the life of the option. We will pay a particular attention to Asian options as they are path-dependent, and one of our goals is to develop path-dependent simulation schemes.

1.2. Mathematical notions

Definition 1.2.1. *Let $(\Omega,\mathscr{F})$ be a measurable space. The function $\mathbb{P}:\mathscr{F}\to[0,1]$ is a measure of probability if*

(1) $\mathbb{P}(\Omega)=1$.

(2) $\forall A\in\mathscr{F}$, $0\leq\mathbb{P}(A)\leq 1$.

(3) $\forall A_1,A_2,\ldots\in\mathscr{F}$ *such that* $A_i\cap A_j=\phi$ *if* $i\neq j$, $\mathbb{P}\left(\cup_{i\geq 1}A_i\right)=\sum_{i\geq 1}\mathbb{P}(A_i)$.

Definition 1.2.2. *The triple $(\Omega,\mathscr{F},\mathbb{P})$ formed of the sample space Ω, a σ-algebra $\mathscr{F}$ on Ω and a probability measure $\mathbb{P}$, is called the probability space.*

We now introduce the notion of stochastic process.

Definition 1.2.3. *Let $(\Omega,\mathscr{F})$ be a measurable space. A stochastic process $X=\{X(t):t\in[0,T]\}$ is a family of random variables, all constructed on the same measurable space.*

Before going any further, we shall define the notion of filtration. This intuitive notion is often used to represent the change in the set of events that can be measured, through gain of information.

Definition 1.2.4. *The filtration* $\mathbb{F} = \{\mathscr{F}_t : t \in [0,T]\}$ *is an increasing collection of* σ*-algebras on a measurable space* $(\Omega, \mathscr{F})$ *such that*

(1) $\forall t \in [0,T],\ \mathscr{F}_t \subseteq \mathscr{F}$.

(2) $\forall t_1, t_2$ *such as* $t_1 \leq t_2$, $\mathscr{F}_{t_1} \subseteq \mathscr{F}_{t_2}$.

Definition 1.2.5. *A stochastic process (see Definition 1.2.3),* $X = \{X(t) : t \in [0,T]\}$*, is adapted with respect to the filtration* $\mathbb{F} = \{\mathscr{F}_t : t \in [0,T]\}$ *if* $\forall t \in [0,T]$, $X(t)$ *is* $\mathscr{F}_t$*-measurable.*

Definition 1.2.6. *The filtration* $\mathbb{F} = \{\mathscr{F}_t : t \in [0,T]\}$ *is the natural filtration with respect to a stochastic process* $X = \{X(t) : t \in [0,T]\}$ *if* $\forall t \in [0,T]$, $\mathscr{F}_t = \sigma(\mathscr{V}$ *and* $X(s) : 0 \leq s \leq t)$ *where* $\mathscr{V}$ *is the set that contains all the zero probability events.*

The stochastic processes to be introduced will use the natural filtration defined in Definition 1.2.6. One important stochastic process in finance is called the Wiener process (or Brownian motion).

Definition 1.2.7. *Let* $W = \{W(t) : t \in [0,T]\}$ *be a Brownian motion built on the filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$*. Thus,* W *have these properties:*

(1) $W(0) = 0$.

(2) $W(t_2) - W(t_1)$ *is independent of* $W(t_4) - W(t_3)$, $\forall 0 < t_1 < t_2 \leq t_3 < t_4 < T$.

(3) $\forall t_1, t_2 \in [0,T]$ *such that* $0 < t_1 < t_2$*, we have that* $W(t_2) - W(t_1)$ *is normally distributed,* $\mathscr{N}(0, t_2 - t_1)$.

(4) The trajectories of $W(t), t \in [0,T]$ *are almost surely continuous.*

The final type of process defined in this preliminary section is the Itô process, frequently used in finance.

Definition 1.2.8. *Let* $I = \{I(t) : t \in [0,T]\}$ *be an Itô process. Then,* I*, defined on the filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$*, is a stochastic process which can be expressed as the sum of an integral with respect to a Wiener process,* W*, and an integral with respect to time*

$$I(t) = I(0) + \int_0^t \alpha(I(s), s)\ ds + \int_0^t \beta(I(s), s)\ dW(s), \tag{1.2.1}$$

where $\alpha(I(s),s)$ *is called the drift term and* $\beta(I(s),s)$ *is the diffusion term. Moreover,* $\alpha(I(s),s)$ *and* $\beta(I(s),s)$ *are adapted processes with respect to the filtration* $\mathbb{F}$ *which satisfy* $\mathbb{P}\left(\int_0^T |\alpha(I(s),s)|ds < \infty\right) = 1$ *and* $\mathbb{P}\left(\int_0^T |\beta(I(s),s)|^2 ds < \infty\right) = 1$.

The following lemma is useful to derive processes of derivative securities (such as call options).

Lemma 1.2.1. (Unidimensional Itô's Lemma). *Let* $f(I(t),t)$ *be the value at time* t *of any derivative security on* X*, where*

$$dI(t) = \alpha(I(t),t)dt + \beta(I(t),t)dW(t). \tag{1.2.2}$$

Thus, $f(X(t),t)$ *follows a diffusion process*

$$df(X(t),t) = \left(\alpha(X(t),t)\frac{\partial f}{\partial X(t)} + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial X^2(t)}\beta^2(X(t),t)\right)dt + \frac{\partial f}{\partial X(t)}\beta(X(t),t)dW(t). \tag{1.2.3}$$

Proof. A formal proof of Itô's lemma requires taking the limit of a sequence of random variables, and can be found in McKean (1969). However, we give an informal derivation of this lemma.

Assume the Itô process (see Definition 1.2.8) is of the form

$$dX(t) = \alpha(X(t),t)dt + \beta(X(t),t)dW(t), \tag{1.2.4}$$

where $W(t)$ is a Wiener process. Expanding $f(X(t),t)$ in a Taylor series in x and t, we have

$$df(X(t),t) = \frac{\partial f}{\partial X(t)}dX(t) + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial X^2(t)}dX^2(t) + ... \tag{1.2.5}$$

and substituting $\alpha(X(t),t)dt + \beta(X(t),t)dW(t)$ for $dX(t)$ gives

$$df(X(t),t) = \frac{\partial f}{\partial X(t)}(\alpha(X(t),t)dt + \beta(X(t),t)dW(t) + \frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial X^2(t)}\left(\alpha^2(X(t),t)dt^2 + 2\alpha(X(t),t)\beta(X(t),t)dtdW(t) + \beta^2(X(t),t)dW^2(t)\right) + ... \tag{1.2.6}$$

In the limit, the dt term tends to 0, the dt^2 and $dtdW(t)$ terms disappear but the $dW^2(t)$ term tends to dt. The latter can be shown if we prove that

$$dW^2 \to \mathbb{E}(dW^2(t)) \text{ since } \mathbb{E}(dW^2(t)) = dt. \tag{1.2.7}$$

By deleting the dt^2 and $dtdW(t)$ terms, substituting dt for $dW^2(t)$ and grouping the dt and $dW(t)$ terms, we obtain

$$df(X(t),t) = \left(\alpha(X(t),t)\frac{\partial f}{\partial X(t)} + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial X^2(t)}\beta^2(X(t),t)\right)dt + \frac{\partial f}{\partial X(t)}\beta(X(t),t)dW(t) \tag{1.2.8}$$

as required. □

A multidimensional version of Itô's lemma also exists. However, it is beyond the scope of this introduction. We refer the reader to Baz and Chacko (2004) for more detail.

We finally introduce two additional theorems: the Radon-Nikodym Theorem and the Cameron-Martin-Girsanov Theorem. These useful results allow one to make change of measure (i.e. going from physical measures to risk-neutral measures).

Theorem 1.2.1 (Radon-Nikodym Theorem)**.** *Let $\mathbb{P}$ and $\mathbb{Q}$ be two equivalent probability measures on $(\Omega,\mathscr{F})$. Then there exists a $\mathscr{F}_t$-adapted random process $Y = \{Y(t) : t \in [0,T]\}$, martingale and starting at $Y(0)$, such that $\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}(Y(T)\ \mathbb{I}_A)$. This random process Y is often noted by $Y(t) = \frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathscr{F}_t}$.*

Proof. See Gauthier (2010a). □

Theorem 1.2.2 (Cameron-Martin-Girsanov Theorem)**.** *For every $i \in \{1,2,...,n\}$, let $\gamma^{(i)} = \{\gamma^{(i)}(t) : t \in [0,T]\}$ be a $\mathbb{F}$-previsible process such that*

$$\mathbb{E}^{\mathbb{P}}\left(\exp\left(\frac{1}{2}\int_0^T \left(\gamma^{(i)}(t)\right)^2 dt\right)\right) < \infty. \tag{1.2.9}$$

Then, there exists a measure $\mathbb{Q}$ on the measurable space $(\Omega,\mathbb{F})$ such that

(1) $\mathbb{Q}$ is equivalent to $\mathbb{P}$.

(2) $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left(-\sum_{i=1}^{n}\int_0^T \gamma^{(i)}(t)\,dW^{(i)}(t) - \frac{1}{2}\int_0^T \sum_{i=1}^{n}\left(\gamma^{(i)}(t)\right)^2 dt\right)$, *where* $W^{(i)}(t)$ *is the value of a Brownian motion at time* t. *This is called the Novikov condition.*

(3) For every $i \in \{1,2,...,n\}$, $\tilde{W}^{(i)} = \{\tilde{W}^{(i)}(t) : t \in [0,T]\}$ *defined by* $\tilde{W}^{(i)}(t) = W^{(i)}(t) + \int_0^t \gamma^{(i)}(s)ds$ *is a Brownian motion on the filtered probability space* $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{Q})$.

PROOF. We refer the reader to Baxter and Rennie (1996) for the proof of Theorem 1.2.2. □

Chapter 2

THE HESTON MODEL

The main objective of this master thesis is to introduce new simulation schemes for the Heston model. However, before introducing simulation methods, we should portray the model itself. We first introduce the Black-Scholes-Merton (BSM) framework in order to grasp the essential facts about Heston (1993). The basic assumptions of the BSM framework as well as the price of European options are provided. We also discuss the shortfalls of the BSM framework. The Heston model is then introduced. The SDEs, PDE, European option prices, asset's c.d.f., and c.d.f. of the integrated variance over time are derived. Finally, we discuss the calibration of the Heston model, which might be challenging.

2.1. Black-Scholes-Merton Framework

In the early 1970s, Black and Scholes (1973) and Merton (1973) achieved a major breakthrough in the pricing and hedging of stock options. Merton and Scholes even received the prestigious Nobel prize in Economics in 1997 (Black died in 1995, thus he was ineligible for the prize). They derived a theoretical value for option prices based on portfolio replication arguments. We now describe their results along with the associated framework. Black and Scholes, and Merton had to set some assumptions:

(1) The underlying asset price $S(t)$ behaves according to a geometric Brownian motion with a constant drift μ and a constant volatility σ under the real probability measure $\mathbb{P}$.

(2) There is no arbitrage opportunity (see Definition 1.1.1).

(3) Transactions occur in continuous time.

(4) It is possible to borrow and lend cash at a known constant risk-free interest rate r.

(5) It is possible to buy and sell any amount, even fractional, of the underlying asset.

(6) There is no transaction cost (this is equivalent to the so-called frictionless market condition) and the underlying security does not pay a dividend.

Definition 2.1.1. *The diffusion process describing the behavior of the underlying asset,* $S = \{S(t), t \geq 0\}$*, in the BSM framework under the real probability measure* $\mathbb{P}$ *is given by*

$$\begin{cases} dS(t) = \mu S(t)dt + \sigma S(t)d\tilde{W}(t) \\ S(0) \text{ is the deterministic initial condition of the SDE} \end{cases} \tag{2.1.1}$$

where

- μ *is the asset growth rate,*
- σ *is the constant volatility of the underlying asset, and*
- $\tilde{W}(t)$ *is the value at time* t *of a Wiener process under the measure* $\mathbb{P}$*.*

$S(t)$ is also known as a geometric Brownian motion.

Proposition 2.1.1. *Under the BSM framework, the European call option price at time* t *is given by*

$$C(S(t), K, T-t, \{\mu, \sigma\}) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \tag{2.1.2}$$

where

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{2.1.3}$$

$d_2 = d_1 - \sigma\sqrt{T-t}$ *and* $\Phi(\cdot)$ *is the c.d.f. of the standard normal distribution.*

Proof. The proof of Proposition 2.1.1 is provided in Appendix A. □

We now describe the behavior of the underlying asset under the risk-neutral measure $\mathbb{Q}$ of the BSM framwork. This new process shall prove to be helpful to derive prices of more complex derivatives.

Corollary 2.1.1. *The diffusion process describing the behavior of the underlying asset in the BSM framework under the risk neutral measure $\mathbb{Q}$ is given by*

$$\begin{cases} dS(t) = rS(t)dt + \sigma S(t)dW(t) \\ S(0) \text{ is the deterministic initial condition of the SDE} \end{cases} \tag{2.1.4}$$

where

- *r is the constant risk-free rate, and*
- *$W(t)$ is the value at time t of a Wiener process under the measure $\mathbb{Q}$.*

Proof. To show this result, we use the Cameron-Martin-Girsanov Theorem introduced in Chapter 1 (see Theorem 1.2.2). Using (2.1.1), we can yield

$$dS(t) = \mu S(t)dt + \sigma S(t)d\left(\int_0^t \gamma(s)ds + \tilde{W}(t)\right). \tag{2.1.5}$$

Suppose $\gamma = \{\gamma(t) = \frac{\mu - r}{\sigma} : t \in [0,T]\}$. It is a $\mathbb{F}$-previsible process. The Novikov condition,

$$\mathbb{E}^{\mathbb{P}}\left(\exp\left(\frac{1}{2}\int_0^T \left(\frac{\mu - r}{\sigma}\right)^2 dt\right)\right) < \infty, \tag{2.1.6}$$

is satisfied since $\gamma(t)$ is a continuous function $\forall t \in [0,T]$. According to Theorem 1.2.2, there exists a measure $\mathbb{Q}$ (our risk-neutral measure here) such that $W = \{W(t) : t \in [0,T]\}$ is defined by

$$W(t) = \tilde{W}(t) + \int_0^t \gamma(s)ds. \tag{2.1.7}$$

Moreover, $W(t)$ is a Brownian motion on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{Q})$. Hence, the diffusion process under the risk neutral measure $\mathbb{Q}$ is given by (2.1.4). □

Using Itô's lemma (see Lemma 1.2.1), we can obtain the SDE for $\log(S(t))$ when $S(t)$ is described by (2.1.4).

Corollary 2.1.2. *Under the BSM framework and conditional on the current time t, the log-price of the underlying asset at some future time T, $\log(S(T))$, satisfies*

$$\log(S(T)) \overset{d}{=} \mathscr{N}\left(\log S(t) + \left(r - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right) \tag{2.1.8}$$

under the risk-neutral measure $\mathbb{Q}$.

PROOF. Let $G(t) = \log(S(t))$. Since

$$\frac{\partial G(t)}{\partial S(t)} = \frac{1}{S(t)}\ , \quad \frac{\partial^2 G(t)}{\partial S^2(t)} = -\frac{1}{S^2(t)} \text{ and } \quad \frac{\partial G(t)}{\partial t} = 0, \tag{2.1.9}$$

we have from Itô's lemma that

$$dG(t) = \left(r - \frac{\sigma^2}{2}\right) dt + \sigma dW(t). \tag{2.1.10}$$

Since r and σ are constant terms, (2.1.10) is a generalized Wiener process whose drift and variance terms are also constant. A change in $G(t)$ between time t and some future time T is thus normally distributed such that

$$\log(S(T)) - \log(S(t)) \overset{d}{=} \mathcal{N}\left(\left(r - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right). \tag{2.1.11}$$

□

We conclude this section by discussing some shortfalls of the BSM framework.

(1) The returns in the BSM framework are implicitly assumed to follow a normal distribution. However, it is a known fact that the return's true distributions are leptokurtic (i.e. tend to exhibit a greater variability that it would be the case if they were normally distributed) (Hurvich, 2000). Moreover, Taleb (2011) strongly criticizes the use of the normal distribution in finance. He proposed the term *GIF* for the *Great Intellectual Fraud*.

(2) Not everyone is able to borrow and lend cash at a known constant risk-free interest rate r. Only a small percentage of the investors can do so. Thus, this assumption does not seem legitimate.

(3) The volatility is obviously not constant in the market. The empirical distribution of the returns has fat tails and high central peaks. These characteristics are observed in mixtures of distributions with different variances (see Gatheral (2006)). This should motivate us to model the variance as a random variable.

2.2. HESTON MODEL'S ASSUMPTIONS

In order to correct the third criticism of the BSM model, many tried to introduce a new random variable for the volatility (or the variance) of the asset in

their models. With this idea came the inception of the first stochastic volatility models.

Heston (1993) was not the only researcher to work within a stochastic volatility framework. Before 1993, Stein and Stein (1991) and Hull and White (1987) published famous stochastic volatility models based on SDEs. Moreover, others researchers tried a different approach: they modeled stochastic volatility with GARCH processes. One of the first published papers about option pricing field under GARCH is Duan *et al.* (1999). Heston and Nandi (2000) also developed a pricing model using GARCH processes. Heston (1993) was however the first to propose a semi-closed solution for a European option and, to our knowledge, the first to use the inversion of the characteristic function to recover the risk neutral probabilities.

The assumptions of the Heston model are very similar to those introduced in Section 2.1, except for the first assumption. Under Heston's framework, the underlying asset price $S(t)$ behaves according to something close to a geometric Brownian motion with a constant drift μ under the real probability measure $\mathbb{P}$. However, the variance, $V = \{V(t) : t \geq 0\}$, is not constant; it behaves according to the square-root process (we shall come back to this claim in Section 2.7).

2.3. Heston's framework under real probability measure $\mathbb{P}$

Definition 2.3.1. (Heston model, partial definition). *Under the real probability measure $\mathbb{P}$, the Heston model is given by*

$$\begin{cases} dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)d\tilde{W}_S(t) & (2.3.1) \\ d\sqrt{V(t)} = -\beta\sqrt{V(t)}dt + \delta d\tilde{W}_V(t) & (2.3.2) \\ S(0) \text{ and } V(0) \text{ are deterministic initial conditions of the SDEs} \end{cases}$$

where μ is the asset growth rate, $-\beta$ is the drift of the volatility, δ is the volatility of the volatility of $\{\sqrt{V(t)} : t \in [0,T]\}$, $\tilde{W}_S(t)$ is the value of a Brownian motion at time t under $\mathbb{P}$, $\tilde{W}_V(t)$ is the value at time t of a Brownian motion under $\mathbb{P}$ correlated with $\tilde{W}_S(t)$ such that $corr(\tilde{W}_S(t), \tilde{W}_V(t)) = \rho$, $-1 \leq \rho \leq 1$.

Using Itô's lemma (see Lemma 1.2.1), one can find the variance process $V = \{V(t) : t \geq 0\}$ (instead of the volatility process), given by

$$dV(t) = \left(\delta^2 - 2\beta V(t)\right) dt + 2\delta\sqrt{V(t)}d\tilde{W}_V(t). \tag{2.3.3}$$

This can be reexpressed (see Cox *et al.* (1985)) as

$$dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}d\tilde{W}_V(t) \tag{2.3.4}$$

where κ is the speed of the mean-reversion, θ is the average level of the variance process V, and σ is the volatility of the variance process V.

Definition 2.3.2. (Heston model, formal definition). *Under the real probability measure $\mathbb{P}$, the Heston model is given by*

$$\begin{cases} dS(t) = \mu S(t)dt + \sqrt{V(t)}S(t)d\tilde{W}_S(t) & (2.3.5) \\ dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}d\tilde{W}_V(t) & (2.3.6) \\ S(0) \text{ and } V(0) \text{ are deterministic initial conditions of the SDEs} \end{cases}$$

.

We can also derive the process associated to the logarithm of the price, $\log(S(t))$. This process shall be useful for implementing our simulation schemes.

Corollary 2.3.1. (Heston model, alternate definition). *Under the real probability measure $\mathbb{P}$, the Heston model is given by*

$$\begin{cases} dX(t) = \left(\mu - \frac{1}{2}V(t)\right) dt + \sqrt{V(t)}d\tilde{W}_X(t) & (2.3.7) \\ dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}d\tilde{W}_V(t) & (2.3.8) \\ X(0) \text{ and } V(0) \text{ are deterministic initial conditions of the SDEs} \end{cases}$$

where $X(t) = \log(S(t))$ and $\tilde{W}_X(t) = \tilde{W}_S(t)$.

Proof. It is a direct application of Itô's lemma and is thus omitted. □

Note that one can interpret the BSM framework as a special case of the Heston model. The BSM framework is a special case of the Heston framework if $\kappa = \theta = \sigma = \rho = 0$ and $V(s) = V(t)$ is constant $\forall s \in [t, T]$ (and is equal to the variance in the BSM model).

2.4. Derivation of the European call option price under the Heston model

A big achievement in Heston (1993) is the successful derivation of a semi-closed formula for European call option prices.

Proposition 2.4.1. *The price of a European call option under the Heston framework is*

$$C(S(t),K,T-t,\{r,\kappa,\theta,V(t),\sigma,\rho\}) = S(t)P_1(\log(S(t)),V(t),T-t) - Ke^{r(T-t)}P_2(\log(S(t)),V(t),T-t), \quad (2.4.1)$$

where

$$P_j(\log(S(t)),V(t),T-t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-\mathrm{i}\phi\log(K)}f_j(\phi,\log(S(t)),V(t))}{\mathrm{i}\phi}\right) d\phi \quad (2.4.2)$$

for $j = 1,2$*, and* Re *is the real part of a complex number. The characteristic functions* f_1 *and* f_2 *are given by*

$$f_j(\phi,\log(S(t)),V(t),T-t) = \exp\left(C_j(T-t,\phi) + D_j(T-t,\phi)V(t) + \mathrm{i}\phi\log(S(t))\right), \quad (2.4.3)$$

where

$$C_j(T-t,\phi) = r\mathrm{i}\phi(T-t) + \frac{a}{\sigma^2}\left((b_j - \rho\sigma\mathrm{i}\phi + h_j)(T-t) - \log\left(\frac{1-g_je^{h_j(T-t)}}{1-g_j}\right)\right), \quad (2.4.4)$$

$$D_j(T-t,\phi) = \frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{\sigma^2}\left(\frac{1-e^{h_j(T-t)}}{1-g_je^{h_j(T-t)}}\right), \quad (2.4.5)$$

$$g_j = \frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{b_j - \rho\sigma\mathrm{i}\phi - h_j}, \quad (2.4.6)$$

$$h_j = \sqrt{(\rho\sigma\mathrm{i}\phi - b_j)^2 - \sigma^2(2u_j\mathrm{i}\phi - \phi^2)}, \quad (2.4.7)$$

$$b_1 = \kappa - \rho\sigma, \quad b_2 = \kappa, \quad (2.4.8)$$

$$u_1 = \frac{1}{2}, \quad u_2 = -\frac{1}{2}, \tag{2.4.9}$$

and

$$a = \kappa\theta. \tag{2.4.10}$$

PROOF. The proof of Proposition 2.4.1 is provided in Appendix B. □

In order to obtain a numerical value for the price, one could use numerical integration techniques such as Gaussian quadratures. See Appendix C for a discussion about the different methods used for the numerical integration in Proposition 2.4.1.

Albrecher *et al.* (2007) slightly corrected Heston's formula for the European call option. They found a few numerical problems in the equations introduced in Heston (1993). In order to correct them, they multiplied the numerator and denominator of (2.4.5) by $\frac{1}{g_j}e^{-h_j(T-t)}$, which leads to an equivalent form. This new formulation has the advantage of being more stable from a numerical viewpoint.

Corollary 2.4.1. *In Proposition 2.4.1, one could use*

$$D_j(T-t,\phi) = \frac{b_j - \rho\sigma i\phi - h_j}{\sigma^2}\left(\frac{1-e^{-h_j(T-t)}}{1-c_j e^{-h_j(T-t)}}\right), \tag{2.4.11}$$

$$C_j(T-t,\phi) = \\ ri\phi(T-t) + \frac{a}{\sigma^2}\left((b_j - \rho\sigma i\phi - h_j)(T-t) - 2\log\left(\frac{1-c_j e^{-h_j(T-t)}}{1-c_j}\right)\right), \tag{2.4.12}$$

where $c_j = \frac{1}{g_j}$.

By applying these changes, the new C_j *and* D_j *are equivalent to those defined in Proposition 2.4.1.*

PROOF. See Albrecher *et al.* (2007). □

Figure 2.1 illustrates the numerical problem in Heston's formulation along with Albrecher *et al.* (2007)'s correction for this problem. The figure shows the behavior of the real part of the characteristic function f_1 for different values of

ϕ. Even if the f_1 defined by Heston tends to the one derived by Albrecher *et al.* (2007) when ϕ goes to infinity, there could be problems for smaller values of ϕ.

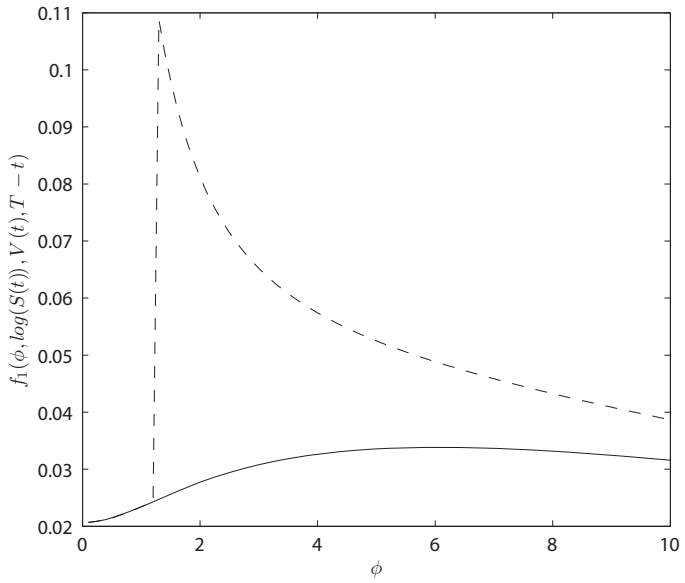

FIGURE 2.1. Plot of $f_1(\phi, \log(S(t)), V(t), T-t)$ against ϕ for Heston (1993)'s formulation (dashed line) and Albrecher *et al.* (2007)'s formulation (solid line). We use $\theta = 0.04$, $r = 0$, $\kappa = 2$, $\sigma = 3$, $T-t = 2$, $\rho = -0.8$, $V(0) = 0.04$ and $S(0) = K = 100$.

2.5. Heston model's SDE under risk-neutral probability measure $\mathbb{Q}$

Corollary 2.5.1. (Heston model, formal definition). *Under the risk-neutral probability measure $\mathbb{Q}$, the Heston model is given by*

$$\begin{cases} dS(t) = rS(t)dt + \sqrt{V(t)}S(t)dW_S(t) & (2.5.1) \\ dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}dW_V(t) & (2.5.2) \\ S(0) \text{ and } V(0) \text{ are deterministic initial conditions of the SDEs} \end{cases}$$

where

- $W_S(t)$ *and* $W_V(t)$ *are the values at time* t *of two Brownian motions under* $\mathbb{Q}$, *and*
- ρ *is the correlation between the Wiener processes,* $corr(W_S(t), W_V(t)) = \rho$.

PROOF. To show this result, we use the Cameron-Martin-Girsanov Theorem introduced in Chapter 1 (see Theorem 1.2.2). Using (2.3.5), we obtain

$$dS(t) = rS(t)dt + \sqrt{V(t)}S(t)d\left(\int_0^t \gamma_1(s)ds + \tilde{W}_S(t)\right). \tag{2.5.3}$$

Let $\gamma_1 = \left\{\gamma_1(t) = \frac{\mu - r}{\sqrt{V(t)}} : t \in [0,T]\right\}$, a $\mathbb{F}$-previsible process such that

$$\mathbb{E}^{\mathbb{P}}\left(\exp\left(\frac{1}{2}\int_0^T \left(\frac{\mu - r}{\sqrt{V(t)}}\right)^2 dt\right)\right) < \infty. \tag{2.5.4}$$

Equation (2.5.4) can be proven; however, it is tedious and thus omitted. Let λ be a variable that is free to vary (since our market is incomplete). Using (2.3.6) and the Cholesky decomposition, we can also write

$$\begin{aligned} dV(t) = (\kappa\theta - \kappa V(t) - \lambda V(t))\,dt + \sigma\rho\sqrt{V(t)}dW_S(t) \\ + \sigma\sqrt{1-\rho^2}\sqrt{V(t)}d\left(\tilde{W}_2(t) + \int_0^t \gamma_2(s)ds\right) \end{aligned} \tag{2.5.5}$$

if

$$d\tilde{W}_V(t) = \rho d\tilde{W}_S(t) + \sqrt{1-\rho^2}d\tilde{W}_2(t) \tag{2.5.6}$$

where $\tilde{W}_2(t)$ is the value at time t of an independent Wiener process under measure $\mathbb{P}$. Let $\gamma_2 = \{\gamma_2(t) : t \in [0,T]\}$ such that

$$\gamma_2(t) = \frac{\lambda\sqrt{V(t)}}{\sigma\sqrt{1-\rho^2}} - \frac{\rho(\mu - r)}{\sqrt{1-\rho^2}\sqrt{V(t)}}. \tag{2.5.7}$$

Again, the Novikov condition for γ_2 is satisfied, but the verification of this condition is tedious. Thus, according to Theorem 1.2.2, there exists a measure $\mathbb{Q}$ (our risk-neutral measure here) such that $W_S = \{W_S(t) : t \in [0,T]\}$ defined by

$$W_S(t) = \tilde{W}_S(t) + \int_0^t \gamma_1(s)ds \tag{2.5.8}$$

is the value of a Brownian motion at time t on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{Q})$. Moreover, $W_V = \{W_V(t) : t \in [0,T]\}$ defined by

$$W_2(t) = \tilde{W}_2(t) + \int_0^t \gamma_2(s)ds \tag{2.5.9}$$

is also the value of a Brownian motion at time t on the filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{Q})$. Hence, the diffusion process under the risk neutral measure $\mathbb{Q}$ is given by

$$dS(t) = rS(t)dt + \sqrt{V(t)}S(t)dW_S(t) \tag{2.5.10}$$

and

$$dV(t) = \kappa\theta - \kappa V(t) - \lambda V(t) + \sigma\rho\sqrt{V(t)}dW_S(t) + \sigma\sqrt{1-\rho^2}\sqrt{V(t)}dW_2(t). \tag{2.5.11}$$

Thus,

$$dV(t) = \kappa^*(\theta^* - V(t))\,dt + \sigma\sqrt{V(t)}dW_V(t), \tag{2.5.12}$$

if

$$dW_V(t) = \rho dW_S(t) + \sqrt{1-\rho^2}dW_2(t) \tag{2.5.13}$$

where

$$\kappa^* = \kappa + \lambda \tag{2.5.14}$$

and

$$\theta^* = \frac{\kappa\theta}{\kappa+\lambda}. \tag{2.5.15}$$

Hence, by redefining κ^* and θ^* in (2.5.11), we get (2.5.2). In order to get a similar notation to that used by Heston (1993), we drop the stars on κ^* and θ^*. From this moment, we use almost exclusively the $\mathbb{Q}$ measure. Thus, there will not be any conflit of notation. □

We can also derive the process associated to the logarithm of the price, $\log(S(t))$, just like we did in Section 2.3.

Corollary 2.5.2. (Heston model, alternate definition). *Under the risk-neutral probability measure $\mathbb{Q}$, the Heston model is given by*

$$\begin{cases} dX(t) = \left(r - \frac{1}{2}V(t)\right)dt + \sqrt{V(t)}dW_X(t) & (2.5.16) \\ dV(t) = \kappa(\theta - V(t))\,dt + \sigma\sqrt{V(t)}dW_V(t) & (2.5.17) \\ X(0) \text{ and } V(0) \text{ are deterministic initial conditions of the SDEs} \end{cases}$$

where $X(t) = \log(S(t))$ and $W_X(t) = W_S(t)$.

PROOF. It is a direct application of Itô's lemma and is thus omitted. □

The rest of this chapter is devoted to a thorough analysis of the processes in (2.5.1) and (2.5.2), and to the derivation of associated results.

2.6. Analysis of the asset dynamics

The SDE of the asset price under the Heston model looks like a geometric Brownian motion. However, the volatility is not constant (by opposition to the BSM framework).

Andersen (2007) derives the first two centered moments of the logarithm of the price under the Heston framework. He simplifies the moments by omiting the risk-free rate (e.g. $r = 0$).

Proposition 2.6.1. *The expected value of* $X(T) = \log(S(T))$ *conditional on* $X(t)$ *and* $V(t)$*, for some future time* $T > t$*, is*

$$\mathbb{E}(X(T)|X(t),V(t)) = X(t) + \frac{1}{2\kappa}(\theta - V(t))\left(1 - e^{-\kappa(T-t)}\right) - \frac{1}{2}\theta(T-t) \quad (2.6.1)$$

and the variance is

$$\mathrm{Var}(X(T)|X(t),V(t)) = \frac{\theta}{8\kappa^3}\Omega_1 + \frac{V(t)}{4\kappa^3}\Omega_2, \quad (2.6.2)$$

where

$$\begin{aligned}\Omega_1 = {} & e^{-2\kappa(T-t)}\sigma^2 + 4e^{-\kappa(T-t)}\left((1+\kappa(T-t))\sigma^2 - 2\rho\kappa\sigma(2+\kappa(T-t)) + 2\kappa^2\right) \\ & + (2\kappa(T-t)-5)\sigma^2 - 8\rho\kappa\sigma(\kappa(T-t)-2) + 8\kappa^2(\kappa(T-t)-1) \quad (2.6.3)\end{aligned}$$

and

$$\begin{aligned}\Omega_2 = {} & -e^{-2\kappa(T-t)}\sigma^2 + 2e^{-\kappa(T-t)} \\ & \left(-\kappa(T-t)\sigma^2 + 2\rho\sigma\kappa(1+\kappa(T-t)) - 2\kappa^2\right) + \sigma^2 - 4\kappa\rho\sigma + 4\kappa^2. \quad (2.6.4)\end{aligned}$$

PROOF. See Andersen (2007). □

2.7. Analysis of the variance dynamics

As mentioned earlier, the process $V = \{V(t) : t \geq 0\}$ is identical to the process applied by Cox *et al.* (1985) to short interest rates. It is called the mean-reverting square root process; this type of process is realistic for modelling variances. Many analytical results for this SDE are known and well-documented. Cox *et al.* (1985) were the first to consider this SDE in a financial application (to our knowledge).

In this section, we now list a few important results about this process; we closely follow Andersen (2007).

Proposition 2.7.1. *Let $F_{\chi^2}(z;\nu,\lambda)$ be the c.d.f. of the non-central chi-square distribution with non-centrality parameter λ and ν degrees of freedom*

$$F_{\chi^2}(z;\nu,\lambda) = e^{-\frac{\lambda}{2}} \sum_{j=0}^{\infty} \frac{\left(\frac{\lambda}{2}\right)^j}{j!2^{\frac{\nu}{2}+j}\Gamma\left(\frac{\nu}{2}+j\right)} \int_0^z x^{\frac{\nu}{2}+j-1} e^{-\frac{x}{2}}\, dx. \tag{2.7.1}$$

Consider the variance process $V = \{V(t) : t \geq 0\}$ defined in (2.5.17). Conditional on $V(t)$, the c.d.f. of $V(T)$ is

$$\mathbb{Q}(V(T) \leq z|V(t)) = F_{\chi^2}\left(\frac{zn(t,T)}{e^{-\kappa(T-t)}}; d, V(t)n(t,T)\right), \tag{2.7.2}$$

where

$$d = \frac{4\kappa\theta}{\sigma^2} \tag{2.7.3}$$

and

$$n(t,T) = \frac{4\kappa e^{-\kappa(T-t)}}{\sigma^2\left(1-e^{-\kappa(T-t)}\right)}, \quad T > t. \tag{2.7.4}$$

Proof. See Andersen (2007). □

The moments of the non-central chi-square distribution might be found in many applied probability or statistics books.

Corollary 2.7.1. *Conditional on $V(t)$, the first two central moments of $V(T)$ are*

$$\mathbb{E}(V(T)|V(t)) = \theta + (V(t)-\theta)e^{-\kappa(T-t)} \tag{2.7.5}$$

and

$$\mathrm{Var}(V(T)|V(t)) = \frac{V(t)\sigma^2 e^{-\kappa(T-t)}}{\kappa}\left(1-e^{-\kappa(T-t)}\right)+\frac{\theta\sigma^2}{2\kappa}\left(1-e^{-\kappa(T-t)}\right)^2. \quad (2.7.6)$$

PROOF. See Andersen (2007). □

Andersen (2007) also computed the covariance between the logarithm of the asset price process and the variance process.

Proposition 2.7.2. *The covariance between $X(T)$, the logarithm of the price at time T, and $V(T)$, the variance at time T, conditional on $X(t)$ and $V(t)$, is given by*

$$\mathrm{Cov}(X(T),V(T)|X(t),V(t)) = \frac{\theta\sigma^2}{2\kappa^2}\Omega_3+\frac{V(t)\sigma^2}{2\kappa^2}\Omega_4, \quad (2.7.7)$$

where

$$\Omega_3 = e^{-2\kappa(T-t)}+2\kappa e^{-\kappa(T-t)}\left((T-t)-\frac{2\rho}{\sigma}(1+\kappa(T-t))\right)+\frac{4\kappa\rho-\sigma}{\sigma} \quad (2.7.8)$$

and

$$\Omega_4 = e^{-\kappa(T-t)}\left(1-\kappa(T-t)+\frac{2\rho\kappa^2(T-t)}{\sigma}\right)-e^{-2\kappa(T-t)}. \quad (2.7.9)$$

PROOF. See Andersen (2007). □

We conclude this section by stating the Feller (1951) condition.

Proposition 2.7.3. *Assume that $V(0)>0$. If $2\kappa\theta \geq \sigma^2$ then the variance process, $V=\{V(t):t\geq 0\}$ can never reach zero. If $2\kappa\theta<\sigma^2$, the origin is accessible and strongly reflecting (likelihood of hitting zero is often quite significant.).*

PROOF. See Feller (1951). □

In Figure 2.2, two curves of $\mathbb{Q}(V(T)\leq z|V(t))$ against z for different sets of parameters illustrate the Feller condition. One can easily see that the non-central chi-square cumulative distribution function is non-zero near the origin if the Feller

condition is violated (solid line); this is however not the case when the condition is satisfied (dashed line).

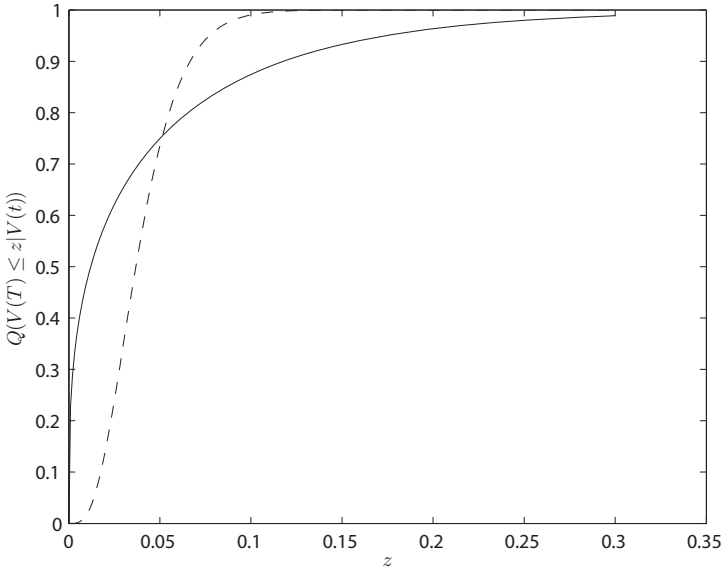

FIGURE 2.2. Plot of $\mathbb{Q}(V(T) \leq z|V(t))$ against z for two specific cases : one where the Feller condition is violated (solid line) and another where it is satisfied (dashed line).

2.8. Hedging portfolio PDE for the Heston model

The hedging portfolio PDE allows us to find prices for various type of derivatives under the measure $\mathbb{Q}$. Moreover, using Heston SDEs along with explicit and implicit assumptions, one can find a unique representation of the model.

To obtain the PDE, we follow the general method for stochastic volatility models (for more detail, see Gatheral (2006)). This method uses a portfolio consisting of the option itself, $f \equiv f(S(t),V(t),t)$, Δ_1 units of the asset, Δ_2 units of another option $u \equiv u(S(t),V(t),t)$ on the same asset, which is used to hedge the volatility, and a money market account. Using this portfolio, we are trying to obtain a hedging strategy which will replicate perfectly the option f.

The value of the portfolio at time t is defined as

$$\Pi(t) = f(S(t),V(t),t) + \Delta_1 S(t) + \Delta_2 u(S(t),V(t),t). \tag{2.8.1}$$

Proposition 2.8.1. *The hedging portfolio PDE for the Heston model using the representation of Corollary 2.5.1 is given by*

$$\frac{\partial f}{\partial t}+\frac{1}{2}V(t)\frac{\partial^2 f}{\partial X^2(t)}+\left(r-\frac{1}{2}V(t)\right)\frac{\partial f}{\partial X(t)}+\rho\sigma V(t)\frac{\partial^2 f}{\partial V(t)\partial X(t)} + \frac{1}{2}\sigma^2 V(t)\frac{\partial^2 f}{\partial V^2(t)}-rf+(\kappa(\theta-V(t))\frac{\partial f}{\partial V(t)}=0 \quad (2.8.2)$$

where $X(t)=\log(S(t))$.

Proof. First of all, we apply Itô's lemma (see Lemma 1.2.1) to f. Thus,

$$df=\frac{\partial f}{\partial t}dt+\frac{\partial f}{\partial S(t)}dS(t)+\frac{\partial f}{\partial V(t)}dV(t)+\frac{1}{2}V(t)S^2(t)\frac{\partial^2 f}{\partial S^2(t)}dt + \frac{1}{2}V(t)\sigma^2\frac{\partial^2 f}{\partial V^2(t)}dt+\sigma V(t)\rho S(t)\frac{\partial^2 f}{\partial S(t)\partial V(t)}dt \quad (2.8.3)$$

since $(dS(t))^2=V(t)S^2(t)dt$, $(dV(t))^2=\sigma^2V(t)dt$, and $dS(t)dV(t)=V(t)S(t)\sigma\rho dt$. Applying Itô's lemma to u yields a similar result:

$$du=\frac{\partial u}{\partial t}dt+\frac{\partial u}{\partial S(t)}dS(t)+\frac{\partial u}{\partial V(t)}dV(t)+\frac{1}{2}V(t)S^2(t)\frac{\partial^2 u}{\partial S^2(t)}dt + \frac{1}{2}V(t)\sigma^2\frac{\partial^2 u}{\partial V^2(t)}dt+\sigma V(t)\rho S(t)\frac{\partial^2 u}{\partial S(t)\partial V(t)}dt \quad (2.8.4)$$

Combining (2.8.3) and (2.8.4), the change in the value of the portfolio at instant t, $d\Pi(t)$, can be written as

$$\begin{aligned} d\Pi(t)=&\left[\frac{\partial f}{\partial t}+\frac{1}{2}V(t)S^2(t)\frac{\partial^2 f}{\partial S^2(t)}+\rho\sigma V(t)S(t)\frac{\partial^2 f}{\partial V(t)\partial S(t)}+\frac{1}{2}\sigma^2V(t)\frac{\partial^2 f}{\partial V(t)}\right]dt \\ &+\Delta_2\left[\frac{\partial u}{\partial t}+\frac{1}{2}V(t)S^2(t)\frac{\partial^2 u}{\partial S^2(t)}+\rho\sigma V(t)S(t)\frac{\partial^2 u}{\partial V(t)\partial S(t)}+\frac{1}{2}\sigma^2V(t)\frac{\partial^2 u}{\partial V(t)}\right]dt \\ &+\left[\frac{\partial f}{\partial S(t)}+\Delta_2\frac{\partial u}{\partial S(t)}+\Delta_1\right]dS(t)+\left[\frac{\partial f}{\partial V(t)}+\Delta_2\frac{\partial u}{\partial V(t)}\right]dV(t). \quad (2.8.5)\end{aligned}$$

In order to hedge against movements in the asset and the variance process, the last two terms in (2.8.5), which involve $dS(t)$ and $dV(t)$, should be zero. This

implies that the proportion in each asset must satisfy

$$\Delta_2 = -\frac{\frac{\partial f}{\partial V(t)}}{\frac{\partial u}{\partial V(t)}} \tag{2.8.6}$$

and

$$\Delta_1 = -\Delta_2 \frac{\partial u}{\partial S(t)} - \frac{\partial f}{\partial S(t)}. \tag{2.8.7}$$

Moreover, the portfolio must earn the risk-free rate, r since $\{e^{rt}S(t) : t \geq 0\}$ behaves like a martingale under the measure $\mathbb{Q}$. Hence,

$$d\Pi(t) = r\Pi(t)dt = r(f + \Delta S(t) + \delta u)dt. \tag{2.8.8}$$

From (2.8.6) and (2.8.7), we can derive the value of the riskless portfolio

$$\begin{aligned} d\Pi(t) = & \\ & \left[\frac{\partial f}{\partial t} + \frac{1}{2}V(t)S^2(t)\frac{\partial^2 f}{\partial S^2(t)} + \rho\sigma V(t)S(t)\frac{\partial^2 f}{\partial V(t)\partial S(t)} + \frac{1}{2}\sigma^2 V(t)\frac{\partial^2 f}{\partial V(t)}\right]dt \\ & + \Delta_2\left[\frac{\partial u}{\partial t} + \frac{1}{2}V(t)S^2(t)\frac{\partial^2 u}{\partial S^2(t)} + \rho\sigma V(t)S(t)\frac{\partial^2 u}{\partial V(t)\partial S(t)} + \frac{1}{2}\sigma^2 V(t)\frac{\partial^2 u}{\partial V(t)}\right]dt, \end{aligned} \tag{2.8.9}$$

which we write as

$$d\Pi(t) = (A + \Delta_2 B)dt \tag{2.8.10}$$

(A and B are implicitly defined). Thus, from (2.8.8),

$$A + \Delta_2 B = r\Pi(t) = r(f + \Delta_1 S(t) + \Delta_2 u), \tag{2.8.11}$$

which yields

$$\frac{A - rf + rS(t)\frac{\partial f}{\partial S(t)}}{\frac{\partial f}{\partial V(t)}} = \frac{B - ru + rS(t)\frac{\partial u}{\partial S(t)}}{\frac{\partial u}{\partial V(t)}}. \tag{2.8.12}$$

The LHS of (2.8.12) only depends on f and the RHS only depends on u. This implies that both sides can be written as a function $y(S(t), V(t), t)$ which corresponds to the term $-\kappa(\theta - V(t))$ in the Heston model (for more details, see Heston (1993)).

By substituting A and rearranging the PDE in (2.8.9), one can get the Heston's PDE expressed in terms of the price $S(t)$

$$\frac{\partial f}{\partial t}+\frac{1}{2}V(t)S^2(t)\frac{\partial^2 f}{\partial S^2(t)}+\rho\sigma V(t)S(t)\frac{\partial^2 f}{\partial V(t)\partial S(t)}+\frac{1}{2}\sigma^2 V(t)\frac{\partial^2 f}{\partial V^2(t)} - rf+rS(t)\frac{\partial f}{\partial S(t)}+[\kappa(\theta-V(t))]\frac{\partial f}{\partial V(t)}=0. \quad (2.8.13)$$

Equation (2.8.13) can be written

$$\frac{\partial f}{\partial t}+\mathscr{A}f-rf=0, \quad (2.8.14)$$

where

$$\mathscr{A}=rS(t)\frac{\partial}{\partial S(t)}+\frac{1}{2}V(t)S^2(t)\frac{\partial^2}{\partial S^2(t)}+[\kappa(\theta-V(t))]\frac{\partial}{\partial V(t)}+ \frac{1}{2}\sigma^2 V(t)\frac{\partial^2}{\partial V^2(t)}+\rho\sigma V(t)S(t)\frac{\partial^2}{\partial V(t)\partial S(t)}. \quad (2.8.15)$$

$\mathscr{A}$ is called the generator of Heston model's SDEs. The first two terms of (2.8.15) correspond to the generator of the BSM model, while the other terms are the corrections for stochastic volatility (for more details, see Lewis (2000)).

Finally, one can find the PDE in terms of the logarithm of the price. Let $X(t)=\log(S(t))$. Thus, the PDE becomes (using Itô's lemma)

$$\frac{\partial f}{\partial t}+\left(r-\frac{1}{2}V(t)\right)\frac{\partial f}{\partial X(t)}+\frac{1}{2}V(t)\frac{\partial^2 f}{\partial X^2(t)}+\rho\sigma V(t)\frac{\partial^2 f}{\partial V(t)\partial X(t)} +\frac{1}{2}\sigma^2 V(t)\frac{\partial^2 f}{\partial V^2(t)}-rf+[\kappa(\theta-V(t))]\frac{\partial f}{\partial V(t)}=0. \quad (2.8.16)$$

□

2.9. Derivation of the c.d.f. of the asset price

Proposition 2.9.1. *Under the Heston model, the c.d.f. of the asset price under the risk neutral probability measure $\mathbb{Q}$ is*

$$\mathbb{Q}(S(T)\le K)=1-\mathbb{Q}(S(T)>K)=1-P_2(\log(S(t)),V(t),T-t)= \frac{1}{2}-\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\phi\log(K)}f_2(\phi,\log(S(t)),V(t))}{i\phi}\right)d\phi, \quad (2.9.1)$$

where the characteristic function f_2 is as in (2.4.3).

PROOF. All of the arguments leading to Proposition 2.9.1 can be found in Appendix B (on the first page). □

2.10. Derivation of the c.d.f. of the integrated variance over time

This integrated variance (over time) formula is a key element for the proposed simulation methods (Chapters 5, 6 and 7).

Proposition 2.10.1. *The c.d.f. of the integrated variance over time under the Heston model, $IV(t,T) = \int_t^T V(s)\ ds$, is given by*

$$\mathbb{Q}(IV(t,T) \leq z) = \frac{2}{\pi}\int_0^\infty \frac{\sin(\phi z)}{\phi}\mathrm{Re}\left(\Xi(\phi)\right)\ d\phi, \tag{2.10.1}$$

where

$$\begin{aligned}\Xi(\phi) = &\frac{\gamma(\phi)e^{-\frac{1}{2}(\gamma(\phi)-\kappa)(T-t)}(1-e^{-\kappa(T-t)})}{\kappa(1-e^{-\gamma(\phi)(T-t)})} \\ &\exp\left(\frac{V(t)+V(T)}{\sigma^2}\left(\frac{\kappa(1+e^{-\kappa(T-t)})}{1-e^{-\kappa(T-t)}} - \frac{\gamma(\phi)(1+e^{-\gamma(\phi)(T-t)})}{1-e^{-\gamma(\phi)(T-t)}}\right)\right) \\ &\mathscr{I}_{\frac{d}{2}-1}\left(\sqrt{V(t)V(T)}\frac{4\gamma(\phi)e^{-\frac{1}{2}\gamma(\phi)(T-t)}}{\sigma^2(1-e^{-\gamma(\phi)(T-t)})}\right) \\ &\mathscr{I}_{\frac{d}{2}-1}\left(\sqrt{V(t)V(T)}\frac{4\kappa e^{-\frac{1}{2}\kappa(T-t)}}{\sigma^2(1-e^{-\kappa(T-t)})}\right),\end{aligned} \tag{2.10.2}$$

$$\gamma(\phi) = \sqrt{\kappa^2 - 2\sigma^2 \mathrm{i}\phi}, \tag{2.10.3}$$

$$d = \frac{4\kappa\theta}{\sigma^2}, \tag{2.10.4}$$

and $\mathscr{I}_\nu(\cdot)$ is the modified Bessel function of the first kind with ν degrees of freedom.

PROOF. See Broadie and Kaya (2006). □

Corollary 2.10.1. *The p.d.f. of the integrated variance over time,* $IV(t,T) = \int_t^T V(s)\,ds$*, under the Heston model is given by*

$$\mathbb{Q}(IV(t,T)=z) = \frac{2}{\pi}\int_0^\infty \cos(\phi z)\mathrm{Re}\left(\Xi(\phi)\right)\,d\phi \tag{2.10.5}$$

Proof. The p.d.f. of a random variable obtained by Lévy's Inversion Theorem. The inversion, using the characteristic function given in (2.10.2), yields

$$f(z) = \mathscr{R}\left(\frac{2}{\pi}\int_0^\infty \cos(\phi z)g(\phi)d\phi\right), \tag{2.10.6}$$

where $f(z)$ is the p.d.f. and $g(\phi)$ is the characteristic function. □

This formulation is not so easy to use, even numerically. In order to simplify this equation, Glasserman and Kim (2011) propose another representation of the integrated variance over time.

Proposition 2.10.2. *The density in (2.10.5) admits the representation*

$$IV(t,T) \overset{d}{=} X_1 + X_2 + \sum_{j=1}^{\eta} Z_j \tag{2.10.7}$$

where X_1*,* X_2*,* η*, and* Z_j $\forall j$ *are mutually independent. The random variables* X_1*,* X_2 *and* Z_j *have the following representations:*

$$X_1 \overset{d}{=} \sum_{n=1}^{\infty}\frac{1}{\gamma_n}\sum_{j=1}^{N_m} A_j, \tag{2.10.8}$$

$$X_2 \overset{d}{=} \sum_{n=1}^{\infty}\frac{1}{\gamma_n}B_n, \tag{2.10.9}$$

$$Z_j \overset{d}{=} \sum_{n=1}^{\infty} C_{n,j}, \tag{2.10.10}$$

where

$$\gamma_n = \frac{\kappa^2(T-t)^2 + 4\pi^2 n^2}{2\sigma^2 t^2}. \tag{2.10.11}$$

Here, A_j *is an exponential random variable with mean 1;* N_m *is a Poisson random variable with mean* $(V(t)+V(T))\lambda_n$ *(independent of the others random variables involved), where*

$$\lambda_n = \frac{16\pi^2 n^2}{\sigma^2(T-t)(\kappa^2(T-t)^2 + 4\pi^2 n^2)}; \tag{2.10.12}$$

B_j is a gamma random variable with a shape parameter of $\frac{d}{2}$ and a scale parameter of 1, where

$$d = \frac{4\kappa\theta}{\sigma^2}. \tag{2.10.13}$$

Moreover, $C_{n,j}$, for all j, is a gamma random variable with a shape parameter of 2 and a scale parameter of 1. Finally, η is a Bessel random variable with parameter

$$z = \frac{2\frac{\kappa}{\sigma^2}}{\sinh\left(\frac{\kappa(T-t)}{2}\right)}\sqrt{V(t)V(T)},\ T \in (t,\infty), \tag{2.10.14}$$

and $\nu = \frac{\delta}{2} - 1$ degrees of freedom. Note that η is a discrete random variable (not to be confused with the Bessel function).

PROOF. See Glasserman and Kim (2011) for the proof of Proposition 2.10.2. □

The moments of this distribution will be useful in our new sampling schemes. Using a gamma expansion and its Laplace transforms, it is very easy to derive the moments of the integrated variance.

Proposition 2.10.3. *Let $C_1 = \coth\left(\frac{\kappa(T-t)}{2}\right)$ and $C_2 = \operatorname{csch}^2\left(\frac{\kappa(T-t)}{2}\right)$. The mean and variance (which is always non-negative) of $IV(t,T)$ are given by*

$$\mathbb{E}(IV(t,T)) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(\eta)E(Z) \tag{2.10.15}$$

and

$$\mathrm{Var}(IV(t,T)) = \sigma_{X_1}^2 + \sigma_{X_2}^2 + \mathbb{E}(\eta)\sigma_Z^2 + (\mathbb{E}(\eta^2) - \mathbb{E}(\eta)^2)\mathbb{E}(Z)^2. \tag{2.10.16}$$

Here,

$$\mathbb{E}(X_1) = (V(t) + V(T))\left(\frac{C_1}{\kappa} - (T-t)\frac{C_2}{2}\right), \tag{2.10.17}$$

$$\sigma_{X_1}^2 = (V(t) + V(T))\left(\sigma^2\frac{C_1}{\kappa^3} + \sigma^2(T-t)\frac{C_2}{2\kappa^2} - \sigma^2(T-t)^2\frac{C_1C_2}{2\kappa}\right), \tag{2.10.18}$$

and

$$\mathbb{E}(X_2) = d\sigma^2\ \frac{-2 + \kappa(T-t)C_1}{4\kappa^2}, \tag{2.10.19}$$

$$\sigma_{X_2}^2 = d\sigma^4\frac{-8 + 2\kappa(T-t)C_1 + \kappa^2(T-t)^2C_2}{8\kappa^4}, \tag{2.10.20}$$

$$\mathbb{E}(Z) = 4\frac{\mathbb{E}(X_2)}{d}, \tag{2.10.21}$$

$$\sigma_Z^2 = 4\frac{\sigma_{X_2}^2}{d}, \tag{2.10.22}$$

with

$$d = \frac{4\kappa\theta}{\sigma^2}. \tag{2.10.23}$$

Furthermore,

$$\mathbb{E}(\eta) = \frac{z\mathscr{I}_{\frac{d}{2}}(z)}{2\mathscr{I}_{\frac{d}{2}-1}(z)}, \tag{2.10.24}$$

and

$$\mathbb{E}(\eta^2) = \frac{z^2\mathscr{I}_{\frac{d}{2}+1}(z)}{4\mathscr{I}_{\frac{d}{2}-1}(z)} + \mathbb{E}(\eta), \tag{2.10.25}$$

where

$$z = \frac{2\frac{\kappa}{\sigma^2}}{\sinh\left(\frac{\kappa(T-t)}{2}\right)}\sqrt{V(t)V(T)} \tag{2.10.26}$$

and $\mathscr{I}_\nu(z)$ is a Bessel of the first kind random variable with ν degrees of freedom evaluated at z. Note that the variances of X_1 and X_2 are always non negative.

PROOF. The mean and variance of X_1, X_2, and Z_j can be calculated using the Laplace transforms introduced in Tse and Wan (2010). We refer the reader to this article for more detail. □

2.11. CALIBRATION TO MARKET DATA: HOW IT COULD BE DONE

Since calibrating the Heston model is not the goal of this thesis, we do not perform a profound analysis of this matter. We only outline the simplest method in the literature. For a review of the most popular methods available, see Moodley (2005).

It is important to draw the line between calibration and estimation. The calibration process is normally used to describe how one would find the best estimates of the parameters according to the information acquired today (implicitly present in today's prices). This gives us the estimates under the risk neutral measure $\mathbb{Q}$ since we take advantage of all the information used to price today's assets and derivatives. The estimation process normally refers to how one would find the best estimates of the parameters using past prices, and the behavior of assets and derivatives prices. This process gives us the estimates under the measure

$\mathbb{P}$. Actually, these processes (estimation and calibration) usually yield different parameters. Thus, one cannot just use empirical estimates for the parameters. For more details on the differences, see Bakshi *et al.* (1997).

The Heston model (under the measure $\mathbb{Q}$) has six parameters that need to be estimated: $\Theta = \{r, \kappa, \theta, \sigma, \rho, V(0)\}$.

The most popular and straightforward approach is to minimize the least squared error or discrepancy between model prices and market prices using call options. More specifically, one can evaluate

$$\hat{\Theta} = \arg\min_{\Theta} \sum_i \omega_i \left[C(S(t), K_i, T_i - t, \Theta) - C^M(S(t), K_i, T_i - t) \right]^{2a}, \tag{2.11.1}$$

where ω_i are weight factors, K_i is the strike price of the ith option, T_i is the maturity of the option, C^M is the market price of an observed option and a is an integer greater than zero (typically, 1). By minimizing the squared error, one shall find $\hat{\Theta}$, the vector of parameters for which the model gives the closest prices to those observed in the market. To find the estimates, we need at least six call option prices.

One easy way to set the weights is to use the inverse of the market's bid-ask spread. Consequently, one can use

$$\omega_i = \frac{1}{C^M_{\text{bid}}(S(t), K_i, T_i - t) - C^M_{\text{ask}}(S(t), K_i, T_i - t)}. \tag{2.11.2}$$

The logic behind this choice is the following: if a call option is liquid, then its bid-ask spread is supposed to be small; if so, its weight is going to be large. Since this call is more liquid, its price is more accurate and we thus put more weight on this observation in our calibration process. This approach is not unique; several other weight formulations exist.

This method is simple; however, several problems emerge from this process:

(1) Finding a global minimum is difficult and strongly depends on the optimization method selected.

(2) Unique solutions to (2.11.1) do not necessarily exist, in which case only local minima can be found.

(3) The solution is dependent on the number of options used, of the strike prices considered, and of the maturity selected.

(4) This method can be very long and is computationally intensive.

Chapter 3

MARKOV CHAIN MONTE CARLO ALGORITHMS

This chapter introduces the two main Markov chain Monte Carlo (MCMC hereafter) algorithms that will be used in this thesis. We start by outlining some important results about Markov chains and MCMC methods.

3.1. Preliminary material on Markov chains

A Markov chain is a mathematical system that undergoes transitions from one state to another in a *chainlike* manner. This random process satisfies the Markovian property.

Definition 3.1.1. *Let $\mathscr{T}$ be a totally ordered index set such that $\mathscr{T} = \{t_0, t_1, ..., t_k\}$. Let $(\chi, \mathscr{F}, \mathbb{P})$ be a triple with a filtration $\{\mathscr{F}_t : t \in \mathscr{T}\}$. Let $X = \{X(t) : t \in \mathscr{T}\}$ be a $\mathscr{F}_t$-adapted stochastic process. X possesses the Markov property if, for each $A \in \chi \subseteq \mathbb{R}$ and $s,t \in \mathscr{T}$ with $s < t$,*

$$\mathbb{P}(X(t) \in A | \mathscr{F}_s) = \mathbb{P}(X(t) \in A | X(s)). \tag{3.1.1}$$

The transitions in a Markov chain are defined by what we call a kernel of transition. We consider the case where the state space is continuous since it is more general and our MCMC schemes shall be defined on continuous state spaces.

Definition 3.1.2. *Let K be a kernel of transition. K is a function defined on $K(x,y) : \chi \times \chi \to [0,1]$ where $\chi \subseteq \mathbb{R}$. It has two main properties:*

(1) $K(x,\cdot)$ is the probability measure for every value $x \in \chi$.

(2) $K(\cdot,A)$ is measurable for every set $A \subseteq \chi$.

The first part means that conditionally on x being the current state, K defines the probability density function of the next state in the chain. The second part says that we can evaluate the probability of the chain jumping into some measurable set A from all possible values x.

Using the last two definitions, we can now define the discrete-time version of a Markov process (on a continuous state space). This definition is partly taken from Revuz (1976).

Definition 3.1.3. *Let $M = \{M(t) : t \in \mathscr{T}\}$ be a discrete-time Markov process on a continuous state space, where $\mathscr{T}$ is an index set. Let $K = \{K_t : t \in \mathscr{T}\}$ be a family of kernels of transition defined on $K_t(x,y) : \chi \times \chi \to [0,1]$ where $\chi \subseteq \mathbb{R}$. M is a Markov process if*

(1) For $t \in \mathscr{T}$ and $x \in \chi$, $K_t(x,\cdot)$ is measurable with $K_t(x,\chi) \leq 1$.

(2) For $t \in \mathscr{T}$ and $A \subseteq \chi$, $K_t(\cdot,A)$ is measurable.

(3) For $s,t \in \mathscr{T}$, $x \in \chi$ and $A \subseteq \chi$,

$$K_{t+s}(x,A) = \int_{\chi} K_s(x,dy)K_t(y,A). \tag{3.1.2}$$

This process possesses the Markov property of Definition 3.1.1. Thus, we can write

$$\mathbb{P}(M(t) \in A|\mathscr{F}_s) = \mathbb{P}(M(t) \in A|M(s)). \tag{3.1.3}$$

3.2. Preliminary material on MCMC

The MCMC schemes are generally used to obtain a sample from a distribution of interest. The idea is to construct a Markov chain on χ which is easily run on a computer, and which has this distribution of interest $f : \mathbb{R}^d \to \mathbb{R}^+$ as a stationary distribution. To ease the notation, let us assume for the rest of this chapter that $d = 1$ and thus that we consider an unidimensional f. We also make an abuse of notation when we refer to this stationary distribution f as our target density. Thus, we want to find a transition kernel $K(x,dy)$, as in Definition 3.1.2, for $x,y \in \chi$, such that

$$\int_{x\in\chi} f(dx)K(x,dy) = f(dy). \tag{3.2.1}$$

If we run a Markov chain for a long time, we expect the distribution of $x^{(n)}$ (the nth element of the Markov chain) to be approximately stationary. However, we do not know if the stationary distribution f is unique for the Markov chain defined by K, and, if it exists, we do not know if the Markov chain converges to its stationary distribution. Thus, in order to have a Markov chain converging to stationarity, some important properties should be satisfied.

In order to obtain a sample from f, we need to satisfy two important properties: we must have a unique stationary distribution f and the Markov chain must converge towards f in distribution.

Even if a Markov chain has a stationary distribution f, it may still fail to converge to its stationary distribution. In order to illustrate this issue, we quote an example from Roberts and Rosenthal (2004):

Example 3.2.1. *Suppose* $\chi = \{1,2,3\}$*, with* $f(\{1\}) = f(\{2\}) = f(\{3\}) = \frac{1}{3}$*. Let* $K(1,\{1\}) = K(1,\{2\}) = K(2,\{1\}) = K(2,\{2\}) = \frac{1}{2}$*, and* $K(3,\{3\}) = 1$*. Then* f *is stationary. However, if* $x^{(0)} = 1$*, then* $x^{(n)} \in \{1,2\}$ *for all* n*, so* $\mathbb{P}(x^{(n)} = 3) = 0$ *for all* n *and the distribution of* $x^{(n)}$ *does not converge to* f*. In fact, here the stationary distribution is not unique, and the distribution of* $x^{(n)}$ *converges to a different stationary distribution defined by* $f(\{1\}) = f(\{2\}) = \frac{1}{2}$*.*

The above example is "reducible". Example 3.2.1 considers the case where χ is countable, and illustrates the classical notion of irreducibility. However, when χ is uncountable, which is the case in this thesis, we instead consider the condition of ϕ-irreducibility.

Definition 3.2.1. *A chain is* ϕ*-irreducible if there exists a non-zero* σ*-finite measure* ϕ *on* χ *such that for all* $A \subseteq \chi$ *with* $\phi(A) > 0$*, and for all* $x \in \chi$*, there exists a positive integer* $n = n(x,A)$ *such that* $K^n(x,A) > 0$ *(where* $K^n(x,A) = \mathbb{P}(x^{(n)} \in A | x^{(0)} = x)$*).*

Even ϕ-irreducible Markov chains might not converge in distribution to f due to periodicity problems. Once again, we quote an example from Roberts and Rosenthal (2004) to illustrate this issue.

Example 3.2.2. *Suppose again* $\chi = \{1,2,3\}$*, with* $f(\{1\}) = f(\{2\}) = f(\{3\}) = \frac{1}{3}$*. Let* $K(1,\{2\}) = K(2,\{3\}) = K(3,\{1\}) = 1$*. Then* f *is stationary, and the chain is*

ϕ-irreducible. However, if $x^{(0)} = 1$, then $x^{(n)} = 1$ whenever n is a multiple of 3, so $\mathbb{P}(x^{(n)} = 1)$ oscillates between 0 and 1. There is again no convergence to f.

To avert this problem, we require aperiodicity of our kernel K.

Definition 3.2.2. *A Markov chain with stationary distribution f is aperiodic if there do not exist $d \geq 2$ and disjoint subsets $\chi_1, \chi_2, ..., \chi_d \subseteq \chi$ with $K(x, \chi_{i+1}) = 1$ for all $x \in \chi_i$ $(1 \leq i \leq d-1)$, and $K(x, \chi_1) = 1$ for all $x \in \chi_d$, such that $f(\chi_1) > 0$ (and hence $f(\chi_i) > 0$ for all i).*

To be assured that the transition kernel possesses a unique stationary distribution and that the Markov chain converges towards this stationary distribution, we need two important properties to be satisfied: ϕ-irreducibility and aperiodicity.

These properties bring us to a result about the convergence of a Markov chain towards its stationary distribution.

Theorem 3.2.1. *If a Markov chain on a state space with countably generated σ-algebra is ϕ-irreducible and aperiodic, and has a stationary distribution f, then for f-almost surely, $x \in \chi$,*

$$\lim_{n \to \infty} ||K^n(x, \cdot) - f(\cdot)|| = 0 \tag{3.2.2}$$

where

$$||\nu_1(\cdot) - \nu_2(\cdot)|| = \sup_{A \in \chi} \|\nu_1(A) - \nu_2(A)\|. \tag{3.2.3}$$

In particular, $\lim_{n \to \infty} K^n(x, A) = f(A)$ for all measurable $A \subseteq \chi$.

Proof. See Roberts and Rosenthal (2004). □

With this final theorem, we are now able to build a Markov chain with a stationary distribution f converging to stationarity. However, we do not know how to construct it yet.

We now introduce the reversibility notion that will help us construct a Markov chain with a given stationary distribution f.

Definition 3.2.3. *A Markov chain on a state space χ is reversible with respect to a probability distribution f on χ, if*

$$f(dx)K(x, dy) = f(dy)K(y, dx), \quad x, y \in \chi. \tag{3.2.4}$$

Proposition 3.2.1. *If a Markov chain is reversible with respect to f, then f is stationary for the chain.*

PROOF.

$$\int_{x\in\chi} f(dx)K(x,dy) = \int_{x\in\chi} f(dy)K(y,dx) = f(dy)\int_{x\in\chi} K(y,dx) = f(dy). \quad (3.2.5)$$

□

Thus, if our Markov chain is reversible with respect to f, then a stationary distribution exists and this stationary distribution is f.

Using the basics of this section, we can implement many MCMC schemes such as the Metropolis-Hastings algorithms and Gibbs samplers.

3.3. Metropolis-Hastings algorithm

The Metropolis algorithm was introduced in Metropolis *et al.* (1953) to sample from the Boltzmann distribution, which is very popular in physics and chemistry. The goal of the method is to generate a Markov chain converging towards a desired density, f. To implement this algorithm, one has to select a proposal density, $q(\cdot|\cdot)$ that is used to produce candidates for the Markov chain. Since f usually is a complicated density, we seek a density q that is easy to simulate from and symmetrical (i.e. $q(x|y) = q(y|x)$). The candidates are then accepted in the sample according to some acceptance rule.

The Metropolis-Hastings algorithm (see Hastings (1970)) is a generalization of the Metropolis algorithm in which the proposal density $q(\cdot|\cdot)$ is not required to be symmetrical.

Algorithm 3.3.1. *Metropolis-Hastings algorithm:*

(1) Initialize $x^{(0)}$, the first element of the chain.

(2) $i \leftarrow 0$.

(3) Generate a candidate $y^{(i+1)}$ from the proposal density $q(\cdot|x^{(i)})$.

(4) Compute the acceptance probability

$$\alpha(x^{(i)}, y^{(i+1)}) = \min\left\{1, \frac{f(y^{(i+1)})}{f(x^{(i)})}\frac{q(x^{(i)}|y^{(i+1)})}{q(y^{(i+1)}|x^{(i)})}\right\}. \quad (3.3.1)$$

(5) Accept $y^{(i+1)}$ *with probability* $\alpha(x^{(i)}, y^{(i+1)})$ *and set*

$$x^{(i+1)} = \begin{cases} y^{(i+1)} & \text{with probability } \alpha(x^{(i)}, y^{(i+1)}) \\ x^{(i)} & \text{otherwise} \end{cases} . \tag{3.3.2}$$

(6) Increment i *to* $i \leftarrow i+1$ *and go back to Step 3.*

The acceptance probability α is the key element of the Metropolis-Hastings algorithm. The densities q and f are not identical and thus we do not want to keep every generated number from $q(\cdot|\cdot)$. After all, we want to obtain a sample from f (and not from q), hence we must accept in our sample values that are representative of the density f only.

Moreover, we shall say that we only keep a certain part of our chain; the first numbers generated are quite dependent on the first value of the chain, $x^{(0)}$. Thus, we shall discard the first generated numbers from our final sample. This is called the "burn-in" period. This "burn-in" period is essentially dependent on the initial element of the chain, q, and f.

3.4. Gibbs sampling

To use the basic form of this algorithm, we must be interested in sampling from some multidimensional probability density function $f : \mathbb{R}^d \to \mathbb{R}^+$, where $d \geq 1$. The exact form of f is not required to implement this algorithm; though, we need to know the conditional distributions of x_i given x_1, x_2, ..., x_{i-1}, x_{i+1}, ..., x_d for all i. We can see the basic Gibbs sampler as a special case of the Metropolis-Hastings algorithm: if $q = f$, then $\alpha = 1$ always.

Algorithm 3.4.1. *Gibbs sampler:*

(1) Initialize $\boldsymbol{x}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, ..., x_d^{(0)}\right)$, *the vector of starting values of the chain.*

(2) $i \leftarrow 0$.

(3) Generate

$$\begin{aligned} x_1^{(i+1)} &\stackrel{d}{=} f(\cdot|x_2^{(i)}, x_3^{(i)}, ..., x_d^{(i)}) \\ x_2^{(i+1)} &\stackrel{d}{=} f(\cdot|x_1^{(i+1)}, x_3^{(i)}, ..., x_d^{(i)}) \\ &\vdots \\ x_d^{(i+1)} &\stackrel{d}{=} f(\cdot|x_1^{(i+1)}, x_2^{(i+1)}, ..., x_{d-1}^{(i+1)}) \end{aligned} \tag{3.4.1}$$

(4) Increment i to $i \leftarrow i+1$ and go back to Step 3.

This scheme is easy to implement, but has some known issues. For example, if two variables are perfectly correlated, the sampler get "stuck" and the Markov chain converges very slowly to the right distribution.

Chapter 4

SIMULATION SCHEMES FOR THE HESTON MODEL: AN EXHAUSTIVE REVIEW

In this chapter, we explain in detail the most popular simulation schemes for the Heston model. Moreover, the pros and cons of each method are provided.

4.1. Euler-Maruyama scheme

The first scheme we describe is the easiest to implement. It is also one of the most popular schemes since it can be used with almost every SDE. It is a generalization of the Euler method for the ODEs.

Definition 4.1.1. *Euler-Marumaya scheme: Let $I(t)$ be an Itô process (see Definition 1.2.8) such that*

$$dI(t) = \alpha(I(t),t)dt + \beta(I(t),t)dW_I(t) \tag{4.1.1}$$

where $I(0)$ is the initial condition. According to the Euler-Marumaya method, the approximation of $I(t)$, denoted $\hat{I}(t)$, is obtained by applying the following steps:

(1) Divide the interval $[0,T]$ into N subintervals of equal length $h = \frac{T}{N}$.

(2) Set $\hat{I}(0) = I(0)$.

(3) Define recursively $\hat{I}(hi)$, $\forall i \in \{1,2,...,N\}$ by

$$\hat{I}(hi) = \hat{I}(h(i-1)) + \alpha\left(I(h(i-1)),h(i-1)\right)h + \beta\left(I(h(i-1)),h(i-1)\right)Z\sqrt{h}, \tag{4.1.2}$$

where Z is a standardized Gaussian random variable with mean zero and variance 1.

Using $\hat{X}$ and $\hat{V}$ to denote discrete-time approximations of X and V respectively the Euler-Maruyama scheme applied to (2.5.16) and (2.5.17) would be

$$\hat{X}(hi) = \hat{X}(h(i-1)) + \left(r - \frac{1}{2}\hat{V}(h(i-1))\right)h + \sqrt{\hat{V}(h(i-1)}Z_X\sqrt{h} \tag{4.1.3}$$

$$\hat{V}(hi) = \hat{V}(h(i-1)) + \kappa(\theta - \hat{V}(h(i-1)))h + \sigma\sqrt{\hat{V}(h(i-1))}Z_V\sqrt{h} \tag{4.1.4}$$

where Z_X and Z_V are standardized Gaussian random variables such that $corr(Z_X, Z_V) = \rho$.

On a computer, one can use this well-known trick based on the Cholesky decomposition to generate correlated Gaussian random variable :

$$Z_V = \Phi^{-1}(U_V), \tag{4.1.5}$$

$$Z_X = \rho Z_V + \sqrt{1-\rho^2}\Phi^{-1}(U_X) \tag{4.1.6}$$

where U_X and U_V are independent uniform (on $[0,1]$) observations and Φ^{-1} is the inverse of the cumulative distribution function of the standardized Gaussian random variable. This last function is present in many numerical computing software (such as MATLAB) or can be computed with Moro (1995)'s algorithm.

As introduced, this scheme is almost impossible to use. When the Feller condition is not satisfied (see Propositon 2.7.3), $\hat{V}(hi)$ could also be negative for some i since the origin is accessible and strongly reflecting since the Euler-Maruyama approximation is linear. However, the variance process cannot be negative; some corrections should thus be made if one wanted to use this scheme.

4.2. A popular fix for the Euler-Maruyama scheme: Lord *et al.* method

As mentioned earlier, when the Feller condition is not satisfied, the Euler-Maruyama scheme has a significant bias: it can generate negative values for $\hat{V}$, which is undesirable. Lord *et al.* (2008) consider an important number of solutions for this problem. They also propose a hotfix for the Euler-Maruyama scheme: the full truncation (FT hereafter) scheme. This method is tailored to minimize the positive bias on European options.

Definition 4.2.1. *Lord* et al. *(2008) fix:*

$$\hat{X}(hi) = \hat{X}(h(i-1)) + \left(r - \frac{1}{2} f(\hat{V}(h(i-1)))\right) h + \sqrt{f(\hat{V}(h)(i-1))} Z_X \sqrt{h} \quad (4.2.1)$$

$$\hat{V}(hi) = \hat{V}(h(i-1)) + \kappa\Big(\theta - f(\hat{V}(h(i-1)))\Big) h + \sigma \sqrt{f(\hat{V}(h(i-1)))} Z_V \sqrt{h} \quad (4.2.2)$$

where $f(x) = \max(0, x)$.

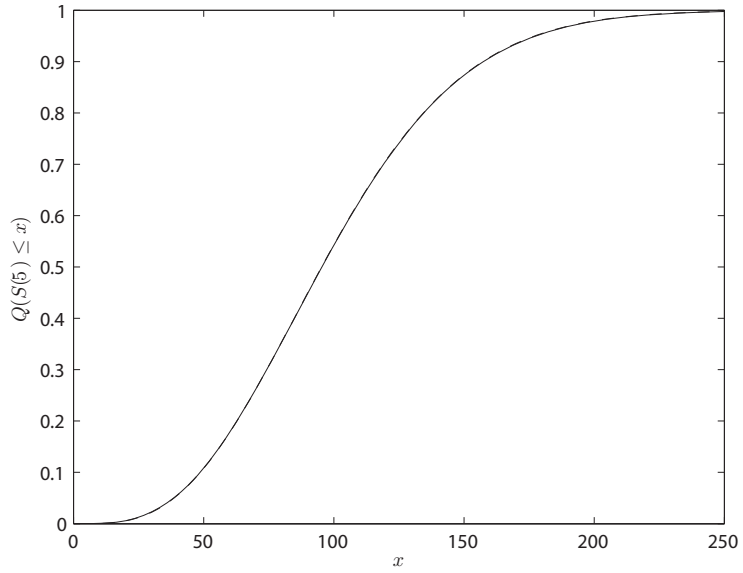

FIGURE 4.1. Example of a simulated c.d.f. using the Euler scheme with the Lord *et al.* (2008) adjustment (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

With this method, the process $\hat{V}$ can have a negative value, at which point it becomes deterministic with an upward drift of $\kappa\theta$. This is the most popular fix in the literature and the one we use hereafter.

Even if this scheme minimizes the positive bias on European call options, it gives poor results when the Feller condition is violated. Figures 4.1 and 4.2 illustrate what happens when this condition is not satisfied.

4.3. MILSTEIN SCHEME

This scheme is very similar to the Euler-Maruyama scheme. It however uses second-order approximations of the SDE in order to increase precision.

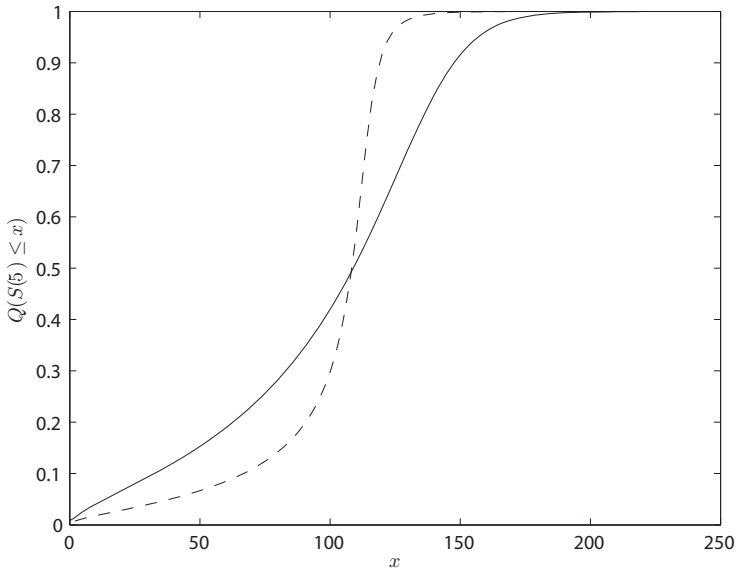

FIGURE 4.2. Example of a simulated c.d.f. using the Euler scheme with the Lord *et al.* (2008) adjustment (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

Definition 4.3.1. *Milstein (1978) scheme: Let $I(t)$ be an Itô process such that*

$$dI(t) = \alpha(I(t),t)dt + \beta(I(t),t)dW_I(t), \tag{4.3.1}$$

where $I(0)$ is the initial condition. According to the Milstein method, the approximation of $I(t)$, denoted $\hat{I}(t)$, is obtained by applying the following steps:

(1) Divide the interval $[0,T]$ into N subintervals of equal length $h = \frac{T}{N}$.

(2) Set $\hat{I}(0) = I(0)$.

(3) Define recursively $\hat{I}(hi)$, $\forall i \in \{1,2,...,N\}$ by

$$\begin{aligned}\hat{I}(hi) = \hat{I}(h(i-1)) &\\ + \alpha\left(I(h(i-1)),h(i-1)\right)h + \beta\left(I(h(i-1)),h(i-1)\right)Z\sqrt{h} &\\ + \frac{1}{2}\beta\left(I(h(i-1)),h(i-1)\right)\beta'\left(I(h(i-1)),h(i-1)\right)\left(hZ^2 - h\right)&\end{aligned} \tag{4.3.2}$$

where Z is a standardized Gaussian random variable with mean zero and variance 1, and $\beta'(I(h(i-1)),h(i-1))$ is the derivative of β with respect to I.

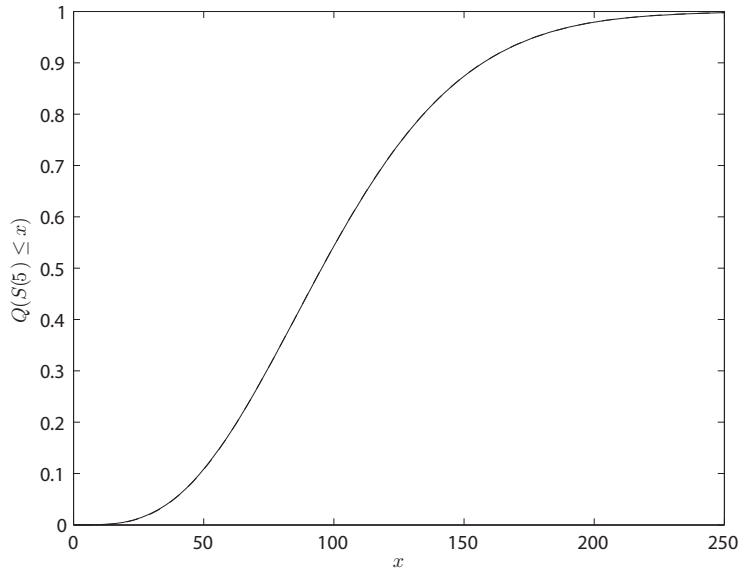

FIGURE 4.3. Example of a simulated c.d.f. using the Milstein (1978) scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

Hence, if we apply Definition 4.3.1 to the process V, we get

$$\hat{V}(hi) = \hat{V}(h(i-1)) + \kappa(\theta - \hat{V}(h(i-1)))h + \sigma\sqrt{\hat{V}(h(i-1))}\sqrt{h}Z_V + \frac{\sigma^2}{4}h(Z_V^2 - 1). \tag{4.3.3}$$

Since the X process is not problematic, we use again the Euler-Maruyama method. Thus,

$$\hat{X}(hi) = \hat{X}(h(i-1)) + \left(r - \frac{1}{2}\hat{V}(h(i-1))\right)h + \sqrt{\hat{V}(h(i-1)}Z_X\sqrt{h} \tag{4.3.4}$$

where Z_X and Z_V are standardized Gaussian random variables such that $corr(Z_X, Z_V) = \rho$.

The Milstein scheme usually converges faster than Euler-Maruyama since the approximation used is to the second order. Thus, Milstein's method gives a better approximation of the behavior of the process V.

It is shown in Gatheral (2006) that $\hat{V}(hi) = 0$ and if $\frac{4\kappa\theta}{\sigma^2} > 1$, then $\hat{V}(h(i+1)) > 0$. If this condition is not satified, then we can also show that the occurrence of negative variance is greatly reduced if we compare to the Euler-Maruyama

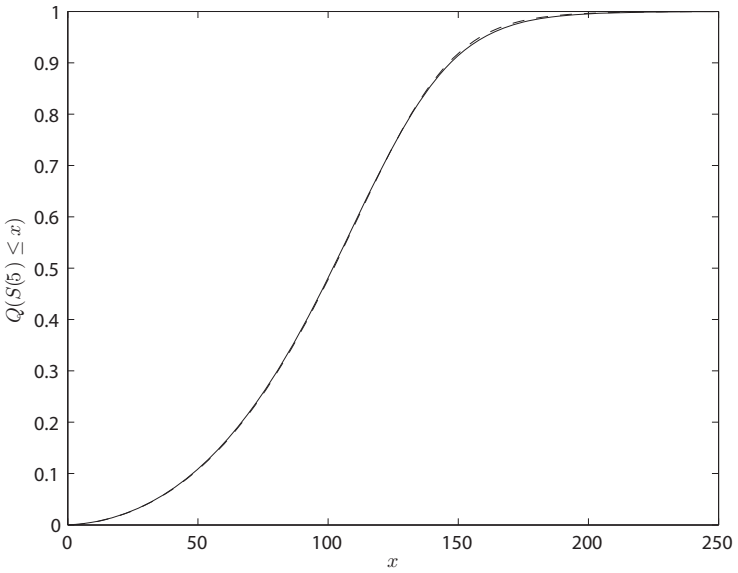

FIGURE 4.4. Example of a simulated c.d.f. using the Milstein (1978) scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated but $\frac{4\kappa\theta}{\sigma^2} > 1$.

scheme. Thus, this algorithm is, in this way, better than the Euler-Maruyama scheme. However, this approximation gives poor results when $\frac{4\kappa\theta}{\sigma^2} \leq 1$. Figures 4.3, 4.4, and 4.5 illustrate this issue. Finally, the Milstein (1978) scheme is not more expensive, computationally speaking, than the Euler-Maruyama scheme. Therefore, we shall always prefer the Milstein scheme to the Euler-Maruyama scheme.

4.4. Broadie and Kaya exact scheme

Broadie and Kaya (2006) propose an exact simulation scheme for the Heston model. Even if this method is elegant and theoretically appealing, its practical implementation is limited. The main issue of this scheme is the lack of speed. Moreover, it is complex. According to Lord *et al.* (2008), simple simulation using the Euler-Maruyama outperforms this scheme in term of computational efficiency.

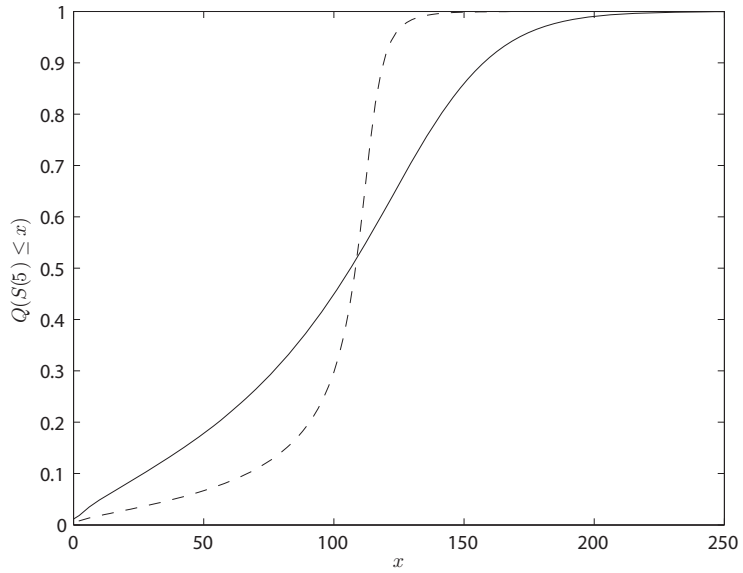

FIGURE 4.5. Example of a simulated c.d.f. using the Milstein (1978) scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when $\frac{4\kappa\theta}{\sigma^2} \leq 1$ (and Feller condition violated).

By using the explicit solution of the asset price,

$$S(hi) = S(h(i-1)) \exp\left(rh - \frac{1}{2}\int_{h(i-1)}^{hi} V(u)\ du + \int_{h(i-1)}^{hi} \sqrt{V(u)}\ dW_S(u) \right), \quad (4.4.1)$$

Itô's lemma, and a Cholesky decomposition, one can obtain

$$\begin{aligned} X(hi) = X(h(i-1)) + rh - \frac{1}{2}\int_{h(i-1)}^{hi} V(u)\ du \\ + \rho \int_{h(i-1)}^{hi} \sqrt{V(u)}\ dW_V(u) + \sqrt{1-\rho^2} \int_{h(i-1)}^{hi} \sqrt{V(u)}\ dW_X(u) \end{aligned} \quad (4.4.2)$$

where $W_X(u)$, the value of a Brownian motion at time u, is independent of $W_V(u)$, the value of another Brownian motion at time u. The integration of the variance process V yields

$$V(hi) = V(h(i-1)) + \int_{h(i-1)}^{hi} \kappa(\theta - V(u))\ du + \sigma \int_{h(i-1)}^{hi} \sqrt{V(u)}\ dW_V(u). \quad (4.4.3)$$

From (4.4.3), one can isolate $\int_{h(i-1)}^{hi}\sqrt{V(u)}\ dW_V(u)$ to get

$$\int_{h(i-1)}^{hi}\sqrt{V(u)}\ dW_V(u)=\sigma^{-1}\Big(V(hi)-V(h(i-1))-\kappa\theta h+\kappa\int_{h(i-1)}^{hi}V(u)\ du\Big). \tag{4.4.4}$$

Then, using (4.4.2), one can obtain

$$\begin{aligned}X(hi)=X(h(i-1))+rh+\frac{\kappa\rho}{\sigma}\int_{h(i-1)}^{hi}V(u)\ du-\frac{1}{2}\int_{h(i-1)}^{hi}V(u)\ du\\ +\frac{\rho}{\sigma}\Big(V(hi)-V(h(i-1))-\kappa\theta h\Big)+\sqrt{1-\rho^2}\int_{h(i-1)}^{hi}\sqrt{V(u)}\ dW_X(u).\end{aligned} \tag{4.4.5}$$

The last equation involves three stochastic quantities which need to be sampled: $V(hi)$ given $V(h(i-1))$, $\int_{h(i-1)}^{hi}V(u)\ du$ given $V(hi)$ and $V(h(i-1))$, and finally, $\int_{h(i-1)}^{hi}\sqrt{V(u)}\ dW_X(u)$ given $\int_{h(i-1)}^{hi}V(u)\ du$. Therefore, the exact scheme can be performed by applying the following steps:

(1) Using the results from Section 2.7, one can generate $\hat{V}(hi)$ given $\hat{V}(h(i-1))$. Broadie and Kaya (2006) use an acceptance and rejection algorithm to perform this step.

(2) Given $\hat{V}(hi)$ and $\hat{V}(h(i-1))$, generate a sample of

$$\widehat{IV}(h(i-1),hi)\equiv\int_{h(i-1)}^{hi}\hat{V}(u)\ du, \tag{4.4.6}$$

which is the integrated variance over time. This step uses numerical inversion since the c.d.f. of this random variable is only known through its characteristic function (see Section 2.10):

$$F(x)=\mathbb{Q}\left(\int_{h(i-1)}^{hi}\hat{V}(u)\ du\le x\right)=\frac{2}{\pi}\int_0^\infty\frac{\sin ux}{u}\mathrm{Re}\left(\Xi(u)\right)\ du \tag{4.4.7}$$

where $\Xi(u)$ is given in (2.10.2). Therefore, from the c.d.f., one could get a sample of $\widehat{IV}(h(i-1),hi)$ using the inversion technique in simulations. Let U be an observation from a uniform (on $[0,1]$) random variable such that

$$F(\widehat{IV}(h(i-1),hi))=\mathbb{Q}\left(\int_{h(i-1)}^{hi}\hat{V}(u)\ du\le\widehat{IV}(h(i-1),hi)\right)=U. \tag{4.4.8}$$

Equation (4.4.8) implies that

$$F^{-1}(U)=\widehat{IV}(h(i-1),hi), \tag{4.4.9}$$

where F^{-1} can be found numerically using the Newton second-order method.

(3) Generate an independent standardized Gaussian random variable Z_X and set:

$$\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\, dW_X(u) = Z_X \sqrt{\widehat{IV}(h(i-1),hi)} \tag{4.4.10}$$

since we know that $\int_{h(i-1)}^{hi} \sqrt{V(u)}\, dW_X(u)$ is normally distributed with mean zero and variance $\widehat{IV}(h(i-1),hi)$.

(4) Given $\hat{X}(h(i-1))$, $\int_{h(i-1)}^{hi} \sqrt{V(u)}\, dW_X(u)$ and $\widehat{IV}(h(i-1),hi)$, use (4.4.5) to compute $\hat{X}(hi)$.

It is suggested in Broadie and Kaya (2006) that one could build a cache for $\widehat{IV}$ for every $\hat{V}(hi)$ and $\hat{V}(h(i-1))$ in order to get better computation time. However, no details are given on this matter.

While this scheme is not directly used in practice, the foundation of Broadie and Kaya (2006) is a starting point for many researchers for simulating stochastic volatility models (and especially, the Heston model).

4.5. Smith almost exact scheme

In order to simplify the application of Broadie and Kaya (2006), Smith (2007) approximates the characteristic function of (2.10.2) using the fact that the geometric and arithmetic means of a sample are near one of the other (the geometric mean, however, is always smaller than the arithmetic mean). Smith (2007) takes a close look at the construction of the characteristic function and discovers that the two parts of (2.10.2), which depend on $\hat{V}(h(i-1))$ and $\hat{V}(hi)$, involve the two means mentionned earlier. Hence, one could try to reduce the number of variables by taking the weighted average of these two means such that

$$z = \omega \frac{\hat{V}(h(i-1)) + \hat{V}(hi)}{2} + (1-\omega)\sqrt{\hat{V}(h(i-1))\hat{V}(hi)} \tag{4.5.1}$$

where $\omega \in [0,1]$. Thus, the approximation of the characteristic function becomes

$$\begin{aligned}\tilde{\Xi}(\phi) = &\frac{\gamma(\phi)e^{-\frac{1}{2}(\gamma(\phi)-\kappa)(T-t)}(1-e^{-\kappa(T-t)})}{\kappa(1-e^{-\gamma(\phi)(T-t)})} \\ &\times \exp\left(\frac{2z}{\sigma^2}\left(\frac{\kappa(1+e^{-\kappa(T-t)})}{1-e^{-\kappa(T-t)}} - \frac{\gamma(\phi)(1+e^{-\gamma(\phi)(T-t)})}{1-e^{-\gamma(\phi)(T-t)}}\right)\right) \\ &\times \mathscr{I}_{\frac{d}{2}-1}\left(z\frac{4\gamma(\phi)e^{-\frac{1}{2}\gamma(\phi)(T-t)}}{\sigma^2(1-e^{-\gamma(\phi)(T-t)})}\right) \\ &\times \mathscr{I}_{\frac{d}{2}-1}\left(z\frac{4\kappa e^{-\frac{1}{2}\kappa(T-t)}}{\sigma^2(1-e^{-\kappa(T-t)})}\right), \quad (4.5.2)\end{aligned}$$

$$\gamma(\phi) = \sqrt{\kappa^2 - 2\sigma^2 \mathrm{i}\phi} \quad (4.5.3)$$

and

$$d = \frac{4\kappa\theta}{\sigma^2} \quad (4.5.4)$$

which only depends on z (and no longer on $\hat{V}(h(i-1))$ and $\hat{V}(hi)$). In our applications, we used $\omega = 0.5$. The rest of Broadie and Kaya (2006) remains unchanged.

Hence, Smith (2007) was able to reduce the number of variables in the characteristic function using a brillant approximation; this approximation is quite useful since it decreases significatively the size of the cache discussed earlier in Section 4.4.

This method is significantly faster than the Broadie and Kaya (2006) scheme. However, it still suffers of long computation times in order to inverse the characteristic function. The results are indeed less acurate than Broadie and Kaya (2006) considering the fact that we approximate the characteristic function.

4.6. Broadie and Kaya drift interpolation scheme

This is another approximation of Broadie and Kaya (2006). Once again, the goal of the Broadie and Kaya drift interpolation scheme (BK-I hereafter) is to speed up the second step of the algorithm presented in Section 4.4.

In order to do that, one could approximate $\widehat{IV}(h(i-1),hi)$ by using a simple quadrature rule called the trapezoidal rule:

$$\widehat{IV}(h(i-1),hi) \approx \frac{(V(h(i-1))+V(hi))h}{2}. \quad (4.6.1)$$

Thus, using this approximation and a small time step, one could get good results in seconds. However, we have to confess that this is a coarse approximation if the time step is not extremely small.

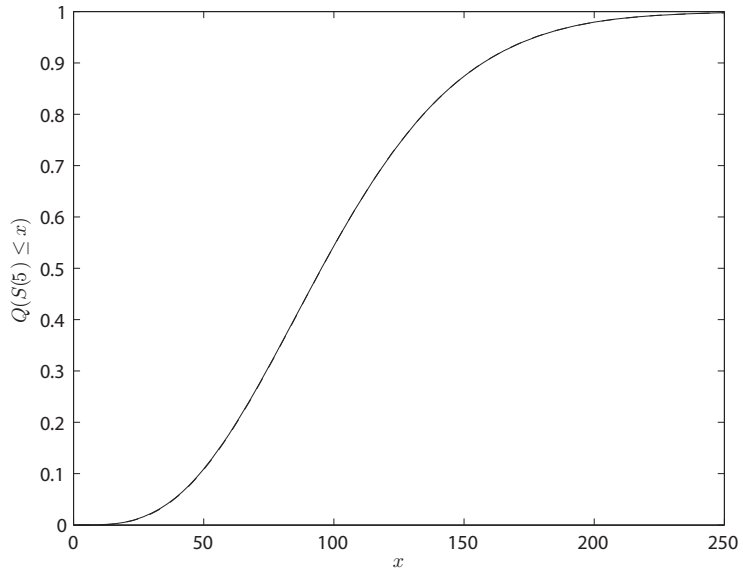

FIGURE 4.6. Example of a simulated c.d.f. using the BK-I scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

Figures 4.6 and 4.7 illustrate two numerical examples of this method. In these examples, we use $h = \frac{1}{32}$, which is very small.

4.7. Andersen's truncated Gaussian scheme

The main idea of this scheme is to sample $\hat{V}(hi)$ from a Gaussian density whose moments were matched. Since the Gaussian density is defined for all $\mathbb{R}$ and the V process cannot be negative, Andersen (2007) rectifies this issue by putting all the probabilities under zero in a Dirac delta function at the origin. This scheme reproduces the asymptotic behavior of $\hat{V}(hi)$ for the large values of $\hat{V}(h(i-1))$. For small values, the density will approximate a chi-square density. In fact, this scheme generates samples that are approximately non-central chi-square distributed.

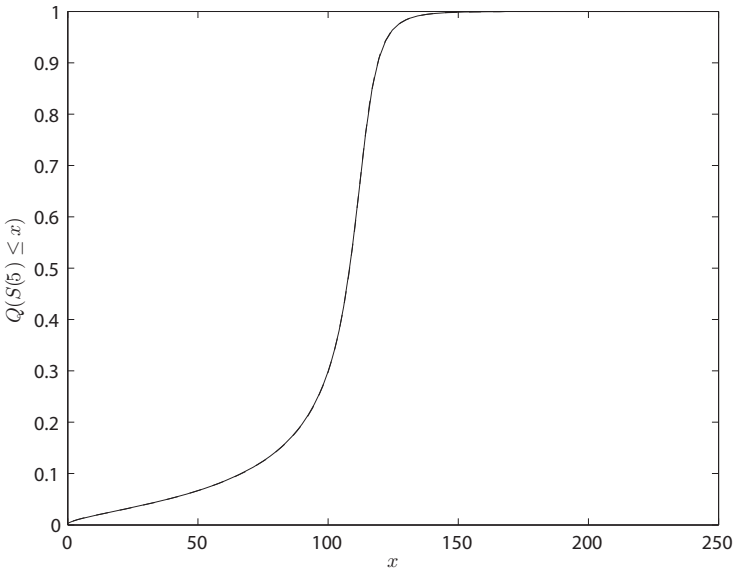

FIGURE 4.7. Example of a simulated c.d.f. using the BK-I scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

We call this scheme truncated Gaussian (thereafter, TG) since the Gaussian density is truncated at the origin. Thus, one can write $\hat{V}(hi)$ in a simpler form:

$$\hat{V}(hi) = \max(0, \mu + \sigma Z_V) \tag{4.7.1}$$

where Z_V is a standardized Gaussian random variable. Moreover, μ and σ are constants that depend on $\hat{V}(h(i-1))$ and on the model's parameters. To find μ and σ, Andersen (2007) uses a moment matching technique: $\mathbb{E}(V(hi))$ and $\mathbb{E}\left(V^2(hi)\right)$ should be matched to $m \equiv \mathbb{E}\left(V(hi)|\hat{V}(h(i-1))\right)$ and $s^2 + m^2 \equiv \mathbb{E}\left(V^2(hi))|\hat{V}(h(i-1))\right)$, which can be calculated using Corollary 2.7.1. Hence, using this technique, we get

$$\mu = \frac{m}{\phi(r(\psi))r^{-1}(\psi) + \Phi(r(\psi))} \tag{4.7.2}$$

and

$$\sigma = \frac{m}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))} \tag{4.7.3}$$

where ϕ is the standardized Gaussian density, Φ is the standardized Gaussian c.d.f., and $r : \mathbb{R} \to \mathbb{R}$ (not to be confused with the risk-free rate) as

$$r(x)\phi(r(x)) + \Phi(r(x))\left(1 + r(x)^2\right) = (1+x)\left(\phi(r(x)) + r(x)\Phi(r(x))\right)^2 \tag{4.7.4}$$

with $\psi = \frac{s^2}{m^2}$.

Due to the non-linear form of (4.7.2) and (4.7.3), the moment-matching exercice cannot be done analytically. Hence, we should rely on numerical methods. However, instead of using a naive brute-force approach, one could define a new variable, $r(x) \equiv r = \frac{\mu}{\sigma}$, and then get

$$\mu = \frac{m}{r^{-1}\phi(r) + \Phi(r)} \tag{4.7.5}$$

and

$$\sigma = \frac{m}{\phi(r) + r\Phi(r)}. \tag{4.7.6}$$

Matching the moments and then simplifying the whole thing yields

$$r\phi(r) + \Phi(r)(1 + r^2) = (1 + \psi)(\phi(r) + r\Phi(r))^2. \tag{4.7.7}$$

Equation (4.7.7) tells us that r is only a function of ψ (i.e. $r = r(\psi)$).

Then, in order to recover the function r, one could use Newton's method and cache the results in order to maximize the speed of the algorithm. One could also cache the values of μ and σ, in respect to ψ, in order to get a quicker algorithm:

$$\mu = f_\mu(\psi)m \tag{4.7.8}$$

and

$$\sigma = f_\sigma(\psi)s \tag{4.7.9}$$

where

$$f_\mu(\psi) = \frac{r(\psi)}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))} \tag{4.7.10}$$

and

$$f_\sigma(\psi) = \frac{\psi^{-1/2}}{\phi(r(\psi)) + r(\psi)\Phi(r(\psi))}. \tag{4.7.11}$$

For the S process, we will use the natural discretization explained in Andersen (2007):

$$\hat{S}(hi) = \exp\Big(\log(\hat{S}(h(i-1))) + rh + K_0 + K_1\hat{V}(h(i-1)) + K_2\hat{V}(hi) + \sqrt{K_3\hat{V}(h(i-1)) + K_4\hat{V}(hi)}Z_S\Big) \quad (4.7.12)$$

with

$$K_0 = -\frac{\rho\kappa\theta}{\sigma}h, \quad (4.7.13)$$

$$K_1 = \gamma_1 h\left(\frac{\kappa\rho}{\sigma} - \frac{1}{2}\right) - \frac{\rho}{\sigma}, \quad (4.7.14)$$

$$K_2 = \gamma_2 h\left(\frac{\kappa\rho}{\sigma} - \frac{1}{2}\right) + \frac{\rho}{\sigma}, \quad (4.7.15)$$

$$K_3 = \gamma_1 h(1-\rho^2), \quad (4.7.16)$$

$$K_4 = \gamma_2 h(1-\rho^2) \quad (4.7.17)$$

and $\gamma_1 = \gamma_2 = \frac{1}{2}$ (using the notation of Andersen (2007).

Algorithm 4.7.1. *The detailed algorithm of the TG scheme:*

(1) Cache $r(\psi)$.

(2) Cache $f_\mu(\psi)$ and $f_\sigma(\psi)$ from the $r(\psi)$ cache.

(3) Compute m and s^2 based on $\hat{V}(h(i-1))$.

(4) Compute $\psi = \frac{s^2}{m^2}$.

(5) Compute μ and σ from $f_\mu(\psi)$ and $f_\sigma(\psi)$ caches.

(6) Draw Z_V from a standardized Gaussian random variable.

(7) Use (4.7.1) to find $\hat{V}(hi)$.

(8) Draw Z_S from a standardized Gaussian random variable independent of Z_V.

(9) Compute $\hat{S}(hi)$ from $\hat{V}(hi)$, $\hat{V}(h(i-1))$, $\hat{S}(h(i-1))$ and Z_S.

Figures 4.8 and 4.9 show two numerical examples that illustrate this method. In general, for the two numerical examples considered, the results seem pretty good. The results of Figure 4.8 (Feller condition satisfied) look less exact than the one from Figure 4.9. This could means that the process V approximation is more biased when the Feller condition is satisfied.

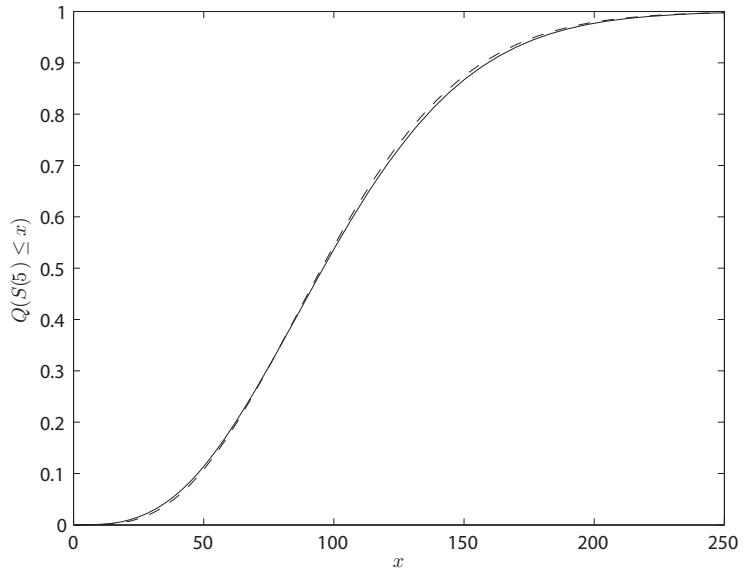

FIGURE 4.8. Example of a simulated c.d.f. using the TG scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

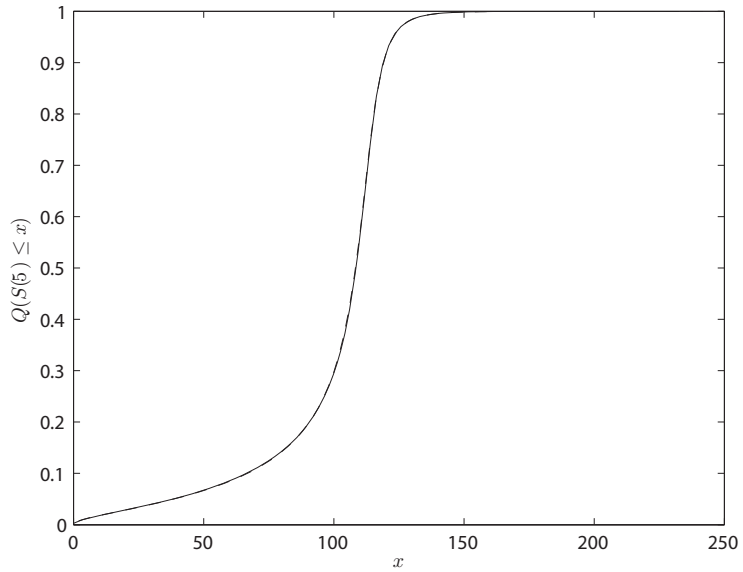

FIGURE 4.9. Example of a simulated c.d.f. using the TG scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

4.8. Andersen's quadratic exponential scheme

This scheme is inspired from the TG scheme considered earlier. In the latter scheme, we saw that the approximated density of $\hat{V}(hi)$ is proportional to $\exp(-x^2/2)$. Though, the real tail decay of $V(hi)$ for the small values of $\hat{V}(h(i-1))$ is less steep than $\exp(-x^2/2)$. Andersen (2007) seeks to correct this behavior in his quadratic exponential (hereafter, QE) scheme.

The first important observation Andersen (2007) made is that a non-central chi-square with a high non-centrality parameter can be well-represented by a power-function applied to a Gaussian variable. While a cubic transformation of a Gaussian variable is preferable, such a scheme would yield negative values (since a cubic transformation is an odd function). Thus, we shall prefer the quadratic representation of Patnaik (1949). Then, for a sufficiently large value of $\hat{V}(h(i-1))$, we get

$$\hat{V}(hi) = a(b + Z_V)^2, \tag{4.8.1}$$

where Z_V is a standardized Gaussian random variable, and a and b are constants determined by moment-matching techniques.

According to Kotz *et al.* (2000), $\hat{V}(hi)$ is distributed as a times a non-central chi-square distribution with one degree of freedom and non-centrality parameter b^2. Then,

$$\mathbb{E}(\hat{V}(hi)) = a(1 + b^2) \tag{4.8.2}$$

and

$$\mathrm{Var}(\hat{V}(hi)) = 2a^2(1 + 2b^2). \tag{4.8.3}$$

Equating (4.8.2) and (4.8.2) to the exact values of m and s^2 introduced in Section 4.7 leads to

$$b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1} \tag{4.8.4}$$

and

$$a = \frac{m}{1 + b^2} \tag{4.8.5}$$

where $\psi = \frac{s^2}{m^2}$ if $\psi \leq 2$.

Equation (4.8.1) does not work well for small $\hat{V}(h(i-1))$. Thus, in order to get a better approximated density for $\hat{V}(hi)$, we set

$$\mathbb{Q}(\hat{V}(hi) \in [x, x+dx]) \approx \left(p\delta(0) + \beta(1-p)e^{-\beta x}\right) dx \tag{4.8.6}$$

where $x \geq 0$, δ is a Dirac delta function, and p and β are non-negative constants to be determined. As in the TG scheme, we have a probability mass at the origin, but this mass p is explicitly specified. Moreover, the mass is supplemented with an exponential tail.

Sampling from (4.8.6) is straightforward since the c.d.f. of the latter equation is easily invertible. Let $\Psi(x)$ be the c.d.f. related to (4.8.6):

$$\Psi(x) = \mathbb{Q}(\hat{V}(hi) \leq x) = p + (1-p)(1-e^{-\beta x}) \tag{4.8.7}$$

for $x \geq 0$. Thus, the inverse is readily computable:

$$\Psi^{-1}(u) = \Psi^{-1}(u, p, \beta) = \begin{cases} 0, & 0 \leq u \leq p \\ \beta^{-1} \log\left(\frac{1-p}{1-u}\right) & p < u \leq 1 \end{cases}. \tag{4.8.8}$$

Using the inversion method, one can sample from (4.8.6).

The constants p and β are derived using, once again, a moment matching method. By direct integration of (4.8.6), it is easy to show that

$$\mathbb{E}(\hat{V}(hi)) = \frac{1-p}{\beta} \tag{4.8.9}$$

and

$$\mathrm{Var}(\hat{V}(hi)) = \frac{1-p^2}{\beta^2}. \tag{4.8.10}$$

Using the exact values computed in Section 4.7, one shall get

$$p = \frac{\psi - 1}{\psi + 1} \tag{4.8.11}$$

and

$$\beta = \frac{1-p}{m}, \tag{4.8.12}$$

where we require $\psi \geq 1$.

We now have two methods to sample $\hat{V}(hi)$, depending on whether $\hat{V}(h(i-1))$ is small or large. Thus, we need to create a switching rule in order to know what scheme to use. According to Andersen, we know that the first part can be used when $\psi \leq 2$ and the second part when $\psi \geq 1$. Fortunately, these domains of

applicability overlap, in a way that at least one of the two schemes can always be used. Thus, one can introduce a critical level $\psi_c \in [1,2]$ as a switching rule.

Finally, for the S process, we will use the strategy explained in Section 4.7.

Algorithm 4.8.1. *The detailed algorithm of the QE scheme :*

(1) Compute m and s^2 from $\hat{V}(h(i-1))$.

(2) Compute $\psi = \frac{s^2}{m^2}$.

(3) If $\psi \leq \psi_C$,

(a) Compute a and b.

(b) Draw Z_V from a standardized Gaussian random variable.

(c) Compute $\hat{V}(hi)) = a(b+Z_V)^2$.

(4) If $\psi > \psi_C$

(a) Compute β and p.

(b) Draw U_V from a uniform random variable, $\mathscr{U}(0,1)$.

(c) Compute $\hat{V}(hi) = \Psi^{-1}(U_V, p, \beta)$.

(5) Draw Z_S from a standardized Gaussian random variable independent of Z_V.

(6) Compute $\hat{S}(hi)$ from $\hat{V}(hi)$, $\hat{V}(h(i-1))$, $\hat{S}(h(i-1))$, and Z_S using 4.7.12.

As witnessed in Figures 4.10 and 4.11, this scheme generally yields good results, even if the Feller condition is violated. However, this method has a big problem called "leaking correlation"; the simulated samples do not exhibit correct correlation behavior according to Zhu (2008).

4.9. Zhu's transformed volatility scheme

The transformed volatility (TV) scheme of Zhu (2008) is based on a transformation of the variance process V. Its goal is simple: it transforms the V process to get the volatility process instead of the variance process used by all the schemes considered up to now. We define $v \equiv \sqrt{V}$. Applying Itô's lemma to v (see Lemma 1.2.1) with (2.5.16) and (2.5.17), one obtains

$$\hat{X}(hi) = \hat{X}(h(i-1)) \left(r - \frac{1}{2}\hat{v}^2(h(i-1)) \right) h + \hat{v}(h(i-1))\sqrt{h} Z_X \tag{4.9.1}$$

$$\hat{v}(hi) = \hat{v}(h(i-1)) + \frac{1}{2}\kappa \Big(\theta_v(h(i-1)) - \hat{v}(h(i-1)) \Big) h + \frac{1}{2}\sigma\sqrt{h} Z_v \tag{4.9.2}$$

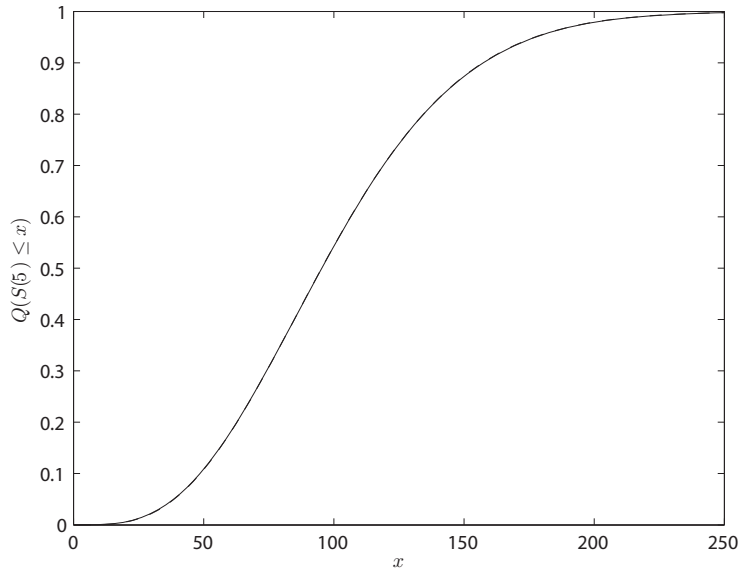

FIGURE 4.10. Example of a simulated c.d.f. using the QE scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

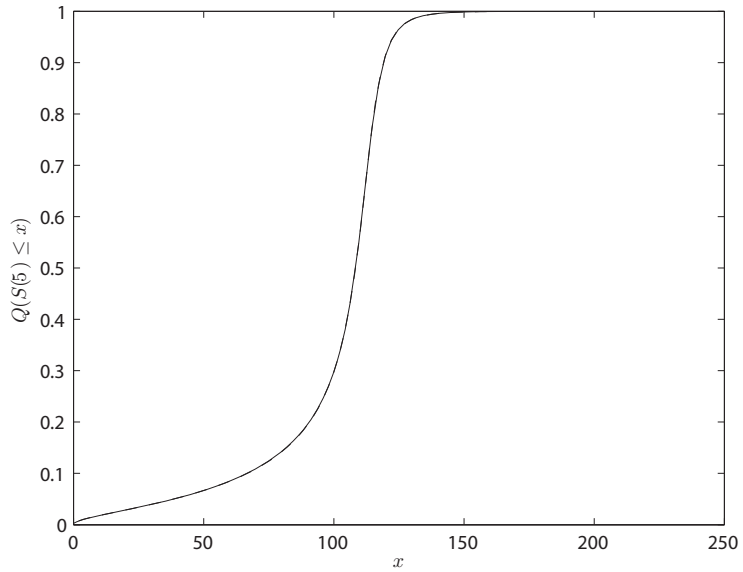

FIGURE 4.11. Example of a simulated c.d.f. using the QE scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

where v is the volatility process, Z_X and Z_v are two correlated standard Gaussian random variables and $\theta_v(h(i-1)) = (\theta - \frac{\sigma^2}{4\kappa})\hat{v}^{-1}(h(i-1))$. Since a naive Euler discretization scheme cannot appropriately capture the erratic behavior of $v^{-1}(t)$, Zhu (2008) proposes two methods to estimate $\theta_v(t)$: one based on a central discretization and the other based on moment matching. Due to the strong oscillation of $\theta_v(t)$ between positive and negative domains, the author admits that the first method fails to recover the average dynamics. However the moment matching technique gives better results. Thus, we have

$$\mathbb{E}(v^2(hi)) = \mathbb{E}(V(hi)) = \ \mathrm{Var}(v(hi)) + \mathbb{E}(v(hi))^2 \tag{4.9.3}$$

where

$$\mathbb{E}(V(hi)) \equiv m_1(h(i-1)) = \theta + (V(h(i-1)) - \theta)e^{-\kappa h}, \tag{4.9.4}$$

$$\mathrm{Var}(v(hi)) \equiv m_2(h(i-1)) = \frac{\sigma^2}{4\kappa}(1 - e^{-\kappa h}) \tag{4.9.5}$$

and

$$\mathbb{E}(v(hi)) = \theta_v + (v(h(i-1)) - \theta_v)e^{-\frac{1}{2}\kappa h}. \tag{4.9.6}$$

Rearranging the above terms yields

$$m_1(h(i-1)) - m_2(h(i-1)) = \theta + (v^2(h(i-1)) - \theta)e^{-\kappa h} - \frac{\sigma^2}{4\kappa}(1 - e^{-\kappa h}). \tag{4.9.7}$$

The single unknown variable is θ_v. Hence, using the discretized process $\hat{v}$ instead of v, one could get

$$\theta_v(hi) = \frac{\beta - \hat{v}(h(i-1))e^{-\frac{1}{2}\kappa h}}{1 - e^{-\frac{1}{2}\kappa h}} \tag{4.9.8}$$

with

$$\beta = \sqrt{\max(0, m_1(h(i-1)) - m_2(h(i-1)))}. \tag{4.9.9}$$

Thus, using (4.9.8) and (4.9.9) coupled with (4.9.1) and (4.9.2), one can easily implement the TV scheme.

According to Zhu (2008), this scheme is very efficient. Also, TV scheme does not have the well-known issue of QE scheme: "leaking correlation".

Figures 4.12 and 4.13 illustrate two numerical examples of this method. As we can see, when the Feller condition is violated, Zhu's method is not as good as the Broadie and Kaya based schemes. Moreover, it is not as good as stated

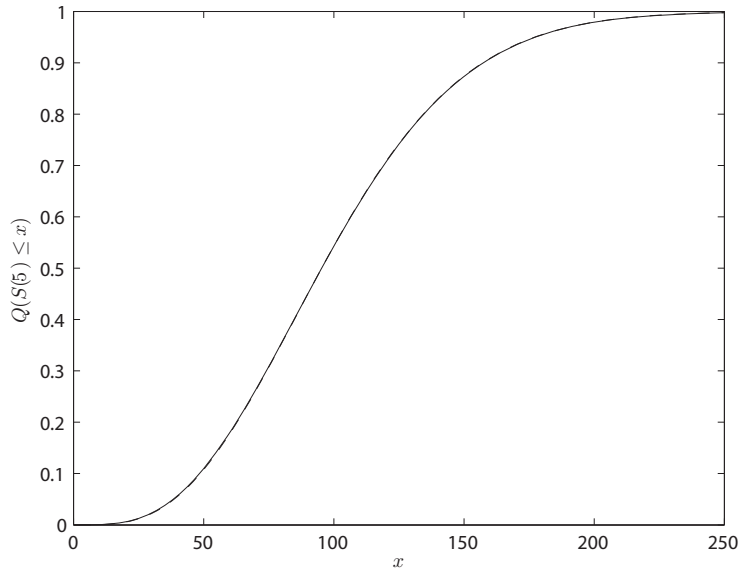

FIGURE 4.12. Example of a simulated c.d.f. using the TV scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

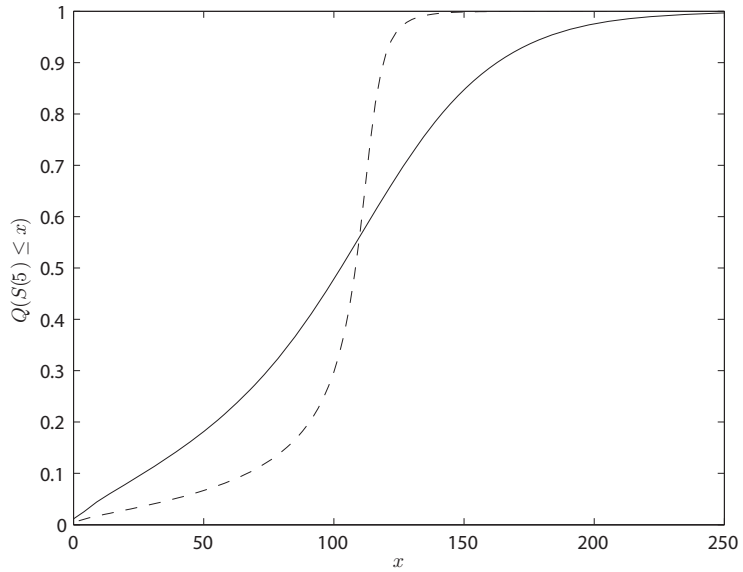

FIGURE 4.13. Example of a simulated c.d.f. using the TV scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

in Zhu (2008). However, when the Feller condition is satisfied, the scheme yields good results.

4.10. Tse and Wan's inverse Gaussian scheme

The inverse Gaussian (thereafter, IG) scheme, which uses the basics of Broadie and Kaya (2006), is based on the fact that the integrated variance over time follows an inverse Gaussian distribution when h goes to infinity.

Proposition 4.10.1. *As $h \to \infty$, the integrated variance approaches in distribution a moment-matched inverse Gaussian distribution.*

Proof. See Tse and Wan (2010) for more detail. □

Using this last proposition, Tse and Wan (2010) thus use the moment-matched inverse Gaussian density instead of the real probability distribution.

Moreover, the simulation from an inverse Gaussian random variable is quite easy. Michael *et al.* (1976) explains the simulation algorithm for an inverse Gaussian random variable; we refer the reader to this article for more details.

Algorithm 4.10.1. *To generate an inverse Gaussian random variable with mean μ and shape parameter λ:*

(1) Generate a random value v from a standardized Gaussian distribution $\mathcal{N}(0,1)$.

(2) Square the value v (i.e. $y = v^2$).

(3) Let:

$$x = \mu + \frac{\mu^2 y}{2\lambda} - \frac{\mu}{2\lambda}\sqrt{4\mu\lambda y + \mu^2 y^2}. \tag{4.10.1}$$

(4) Generate a random variate z from a uniform distribution, $\mathcal{U}(0,1)$.

(5) If $z \leq \frac{\mu}{\mu+x}$, then returns x. Otherwise, returns $\frac{\mu^2}{x}$.

Proof. See Michael *et al.* (1976). □

In a quick numerical experiment, we decided to compare the real p.d.f. to the moment-matched inverse Gaussian. We observe that the inverse Gaussian approximation not only matches very well the overall shape, but also the tails.

Figures 4.14 and 4.15 illustrate two examples of this experiment: the fit is quite impressive. This is why Tse and Wan (2010) scheme is considered "low-bias".

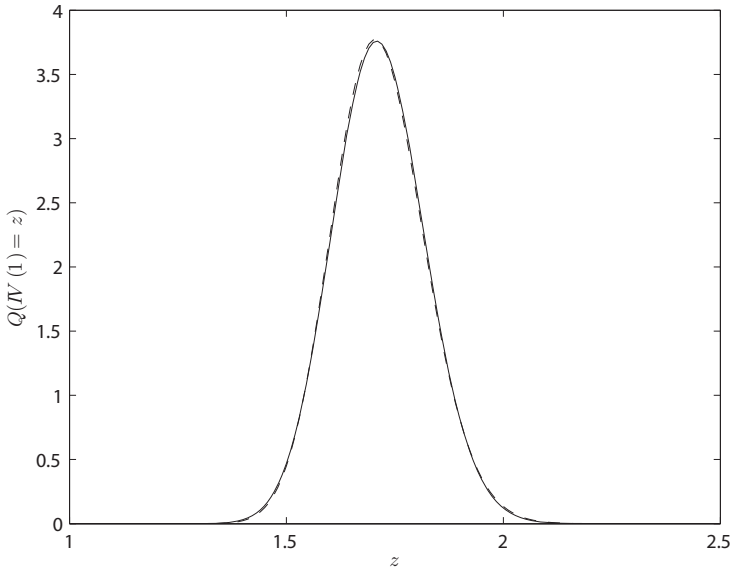

FIGURE 4.14. Comparison of the real p.d.f. of the integrated variance over time (solid line) and the moment-matched inverse Gaussian when the Feller condition is satisfied (dashed line).

They also implemented a new algorithm to sample $V(hi)$ given $V(h(i-1))$. They called it the "IPZ" scheme since they are interpolating from the case when the Poisson variate is equal to zero. This new algorithm is not the main contribution of Tse and Wan (2010), and thus refer the reader to Tse and Wan (2010) to learn more about this algorithm.

Hence, by implementing their new "IPZ" scheme and by approximating the integrated variance by a moment-matched inverse Gaussian, they were able to speed up the exact approach of Broadie and Kaya (2006).

Figures 4.16 and 4.17 illustrate numerical examples with this scheme (without implementing this "IPZ" scheme for $\hat{V}(hi)$). The results, if you compare the real c.d.f. to the simulated one, seem pretty good, even if the Feller condition is violated.

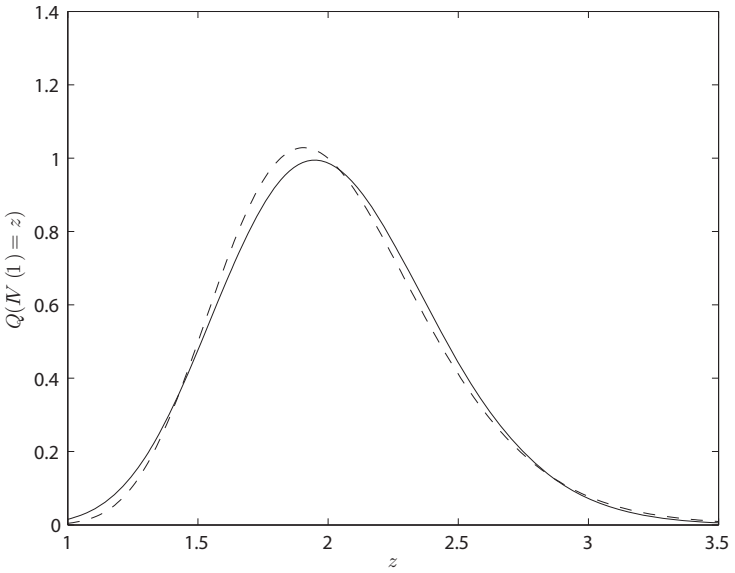

FIGURE 4.15. Comparison of the real p.d.f. of the integrated variance over time (solid line) and the moment-matched inverse Gaussian when the Feller condition is violated (dashed line).

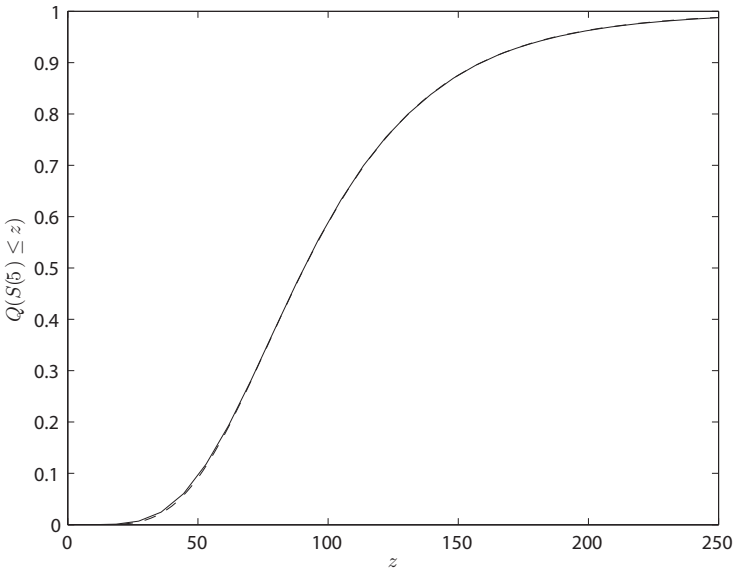

FIGURE 4.16. Example of a simulated c.d.f. using the IG scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

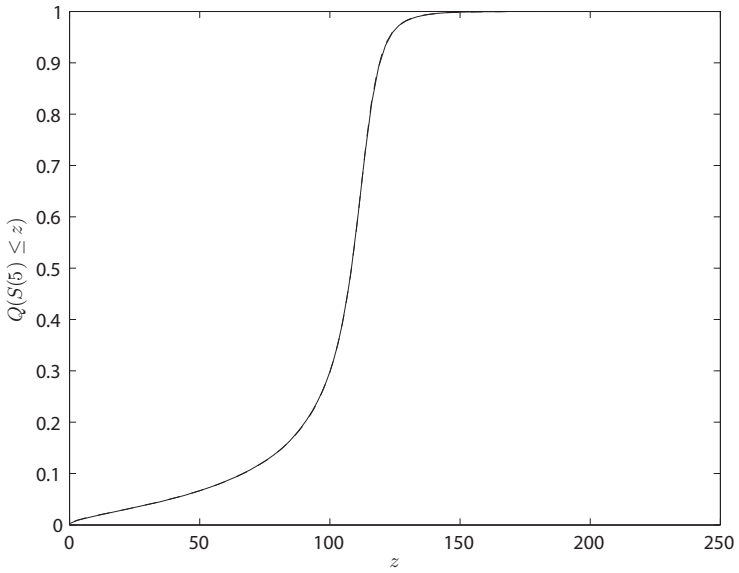

FIGURE 4.17. Example of a simulated c.d.f. using the IG scheme (solid line) versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

4.11. Glasserman and Kim's gamma expansion scheme

This scheme is, again, based on the Broadie and Kaya (2006) scheme. In order to speed up the second step of their algorithm, we use the representation introduced in Proposition 2.10.2.

Instead of inverting the characteristic function, we generate X_1, X_2, and $\sum_{j=1}^{\eta} Z_j$ from Proposition 2.10.2. These random variables are, however, represented by infinite sums. Thus, we can truncate them in order to get a numerical solution in a finite time. This solution will contain a non-negligible bias (depending on K, the number of terms in each approximate sum). Hence, Glasserman and Kim (2011) decided to introduce a correction for this bias.

Proposition 4.11.1. *Let*

$$X_1^K \stackrel{d}{=} \sum_{n=K+1}^{\infty} \frac{1}{\gamma_n} \sum_{j=1}^{N_m} A_j, \tag{4.11.1}$$

$$X_2^K \stackrel{d}{=} \sum_{n=K+1}^{\infty} \frac{1}{\gamma_n} B_n, \tag{4.11.2}$$

$$Z_j^K \stackrel{d}{=} \sum_{n=K+1}^{\infty} C_n, \tag{4.11.3}$$

be the remainder random variables for X_1, X_2 *and* Z_j *(see Proposition 2.10.2). Thus,*

$$\mathbb{E}(X_1^K) = \frac{2(V(hi)+V(h(i-1))h}{\pi^2 K}, \tag{4.11.4}$$

$$\mathrm{Var}(X_1^K) = \frac{2(V(hi)+V(h(i-1))\sigma^2 h^3}{3\pi^4 K^3}, \tag{4.11.5}$$

$$\mathbb{E}(X_2^K) = \frac{\delta\sigma^2 h^2}{4\pi^4 K^3}, \tag{4.11.6}$$

$$\mathrm{Var}(X_2^K) = \frac{\delta\sigma^4 h^4}{24\pi^4 K^3}, \tag{4.11.7}$$

$$\mathbb{E}(Z_j^K) = \frac{\sigma^2 h^2}{\pi^2 K}, \tag{4.11.8}$$

and

$$\mathrm{Var}(Z_j^K) = \frac{\sigma^4 h^4}{6\pi^4 K^3}. \tag{4.11.9}$$

Let R_K *denote any of the remainder random variables* X_1^K, X_2^K *or* Z_j^K. *Then,*

$$\frac{R_K - \mathbb{E}(R_K)}{\sqrt{\mathrm{Var}(R_K)}} \xrightarrow{d} \mathcal{N}(0,1) \tag{4.11.10}$$

as K *goes to infinity.*

Proof. See Glasserman and Kim (2011) for more detail on this proposition. □

Adding these normal random variables should improve the exactness of the truncated series. Hence, with a fixed K, one can generate X_1, X_2 and Z_j by sampling from gamma, exponential, Poisson, and Gaussian random variables such that

$$\hat{X}_1 \stackrel{d}{=} \sum_{n=1}^{K} \frac{1}{\gamma_n} \sum_{j=1}^{N_m} A_j + \mathcal{N}(\mathbb{E}(X_1^K),\ \mathrm{Var}(X_1^K)), \tag{4.11.11}$$

$$\hat{X}_2 \stackrel{d}{=} \sum_{n=K+1}^{K} \frac{1}{\gamma_n} B_n + \mathcal{N}(\mathbb{E}(X_2^K),\ \mathrm{Var}(X_2^K)), \tag{4.11.12}$$

and

$$\hat{Z}_j = \sum_{n=K+1}^{K} C_n + \mathcal{N}(\mathbb{E}(Z_j^K),\ \mathrm{Var}(Z_j^K)) \tag{4.11.13}$$

approximate X_1, X_2 and Z_j when K goes to infinity. Thus, it is easy to efficiently generate the integrated variance (second step of Broadie and Kaya (2006)). The rest of the algorithm is the same as Broadie and Kaya (2006) scheme.

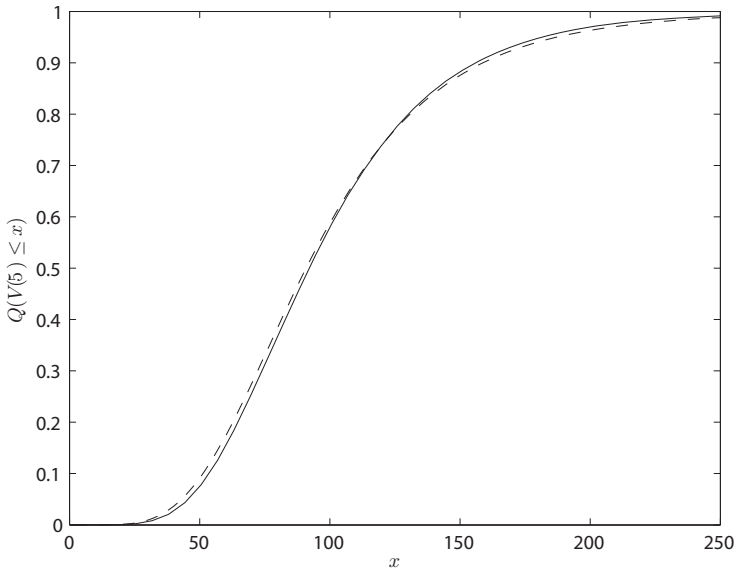

FIGURE 4.18. Example of a simulated c.d.f. using the GE scheme (solid line) with $K = 5$ versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is satisfied.

Figures 4.18 and 4.19 illustrate numerical examples of this scheme. It takes a little more time than the QE scheme according to our tests. Zhu (2008) draw the same conclusion.

4.12. Review of the simulation schemes for the Heston model

In Tables 4.1 et 4.2, we summarize the pros and the cons of the schemes we just explained.

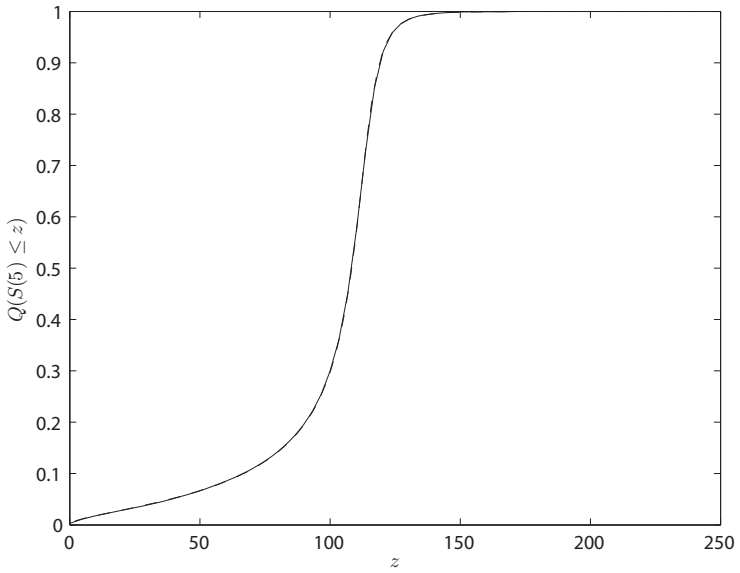

FIGURE 4.19. Example of a simulated c.d.f. using the GE scheme (solid line) with $K = 5$ versus the real c.d.f. using Albrecher *et al.* (2007) formulation (dashed line) when the Feller condition is violated.

TABLE 4.1. Comparison of the most popular simulation schemes for the Heston model.

Name	Pros	Cons
Euler-Maruyama with the Lord *et al.* fix	Fast scheme and good when the Feller condition is satisfied.	Bad when the Feller condition is violated.
Milstein	Better than Euler-Maruyama for the same "computational expenses" and very fast.	Poor results are expected when $\frac{4\lambda\theta}{\sigma^2} \leq 1$.
Broadie and Kaya	Exact solution.	Quite difficult to implement in real life and very long even on a decent computer.

TABLE 4.2. Comparison of the most popular simulation schemes for the Heston model (continued).

Smith	Quicker than the Broadie and Kaya method and almost exact.	Less exact than the Broadie and Kaya scheme and difficult to implement.
Broadie and Kaya with drift interpolation	Quicker and reliable when the time step is small.	Could lead to bad results. Coarse approximation when the time step is large.
Truncated Gaussian	Fast scheme and gives good solutions with a low bias.	The cache requires a numerical inversion which can be long and unstable.
Quadratic exponential	Quick and reliable, even if the Feller condition is violated.	"Leaking correlation" issue.
Transformed volatility	Quick and easy to implement.	Not good when the Feller condition is violated.
Inverse Gaussian	Extremely quick and good results, even when the Feller condition is violated.	This is a low-bias algorithm; thus, the results are not exact.
Gamma expansion	Reliable and quicker than the exact schemes; good even when the Feller condition is violated.	Slower than QE.

Chapter 5

MCMC-BASED SIMULATION ALGORITHM

This chapter presents the details of our first MCMC-based scheme. First of all, we describe the main ideas behind our algorithm. Then, we show how we build the proposal density for the Metropolis-Hastings step. Three efficient caches for the p.d.f. and the first two moments of the integrated variance over time are created. The formal steps of our algorithm are then given. After, we describe how our scheme is tuned and, finally, the new scheme is compared with some of the simulation schemes introduced earlier.

5.1. Intuition behind the new scheme

Before going any further, we shall define the notation used in this chapter:

- $S(t)$ is the asset price at time t and $V(t)$ is the variance of the asset at time t. Thus, $S(0)$ is the initial price of the asset. Moreover, $IV(t,T)$ corresponds to the integrated variance over time from t to T.
- $\hat{S}(t)$ is the simulated price at time t. $\hat{V}$ is the simulated variance at time t and $\widehat{IV}(t,T)$ is the simulated integrated variance over time between t and T.
- κ, θ, ρ, σ, $V(0)$ and r are the Heston model's parameters under $\mathbb{Q}$, which are defined in Section 2.5.1.
- T is the maturity of the derivative. We now divide the interval $[0,T]$ into m equal subintervals of length $h = \frac{T}{m}$.

Our scheme is based on the basics of Broadie and Kaya (2006). However, we decided to speed up the second step (see Section 4.4) by using a Metropolis-Hastings step. By caching the values of the real p.d.f., we are able to efficiently simulate values for the *IV* step. Moreover, we need a good approximation of the *IV* density for our Metropolis-Hastings step. We will use a moment-matched inverse Gaussian (with a slight modification); this is going to be our proposal density. The next two sections discuss these issues.

5.2. Cache implementation

In our scheme, we implement three types of cache: one for the expected value of $IV(h(i-1),hi)$, one for the variance of $IV(h(i-1),hi)$ and one for the p.d.f. of $IV(h(i-1),hi)$.

5.2.1. Expected value and variance of $IV(h(i-1),hi)$

If the moments of $IV(h(i-1),hi)$ were computed directly using Proposition 2.10.3, the modified Bessel function of the first kind would have to be evaluated several times. This procedure is very expensive computationally speaking; hence, caching these values could save time.

The two types of cache are implemented using the same basic ideas since they behave similarly. Tse and Wan (2010) suggested that the first two moments of $IV(h(i-1),hi)$ can be cached efficiently.

Definition 5.2.1. *We define*

$$\mathbb{E}(IV^*(h(i-1),hi)) = \mathbb{E}(X_2) + \mathbb{E}(\eta)E(Z) \tag{5.2.1}$$

and

$$\mathrm{Var}(IV^*(h(i-1),hi)) = \sigma_{X_2}^2 + (\mathbb{E}(\eta^2) - \mathbb{E}(\eta)^2)\mathbb{E}(Z)^2 + \mathbb{E}(\eta)\sigma_Z^2 \tag{5.2.2}$$

using the notation of Propositon 2.10.3.

The two moments of Definition 5.2.1 only depend on the product $V(h(i-1))V(hi)$. Hence, using Proposition 2.10.3, we get $\mathbb{E}(IV(h(i-1),hi)) = \mathbb{E}(IV^*(h(i-1),hi)) + \mathbb{E}(X_1)$ and $\mathrm{Var}(IV(h(i-1),hi)) = \mathrm{Var}(IV^*(h(i-1),hi)) + \sigma_{X_1}^2$. We know that $\mathbb{E}(X_1)$ and $\sigma_{X_1}^2$ depend only on $V(h(i-1)) + V(hi)$, and do not require any

modified Bessel evaluation; thus, it is computationally inexpensive to compute $\mathbb{E}(X_1)$ and $\sigma^2_{X_1}$ at a later moment.

Proposition 5.2.1. *For fast moments calculation, one could use the following steps:*

(1) Precompute $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *using an equally spaced grid on the values of* $V(h(i-1))V(hi)$. *An optimal grid will be introduced in Subsection 5.5.3.*

(2) Compute $\mathbb{E}(X_1)$ *and* $\sigma^2_{X_1}$.

(3) Use linear interpolation to approximate $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *from their respective cache.*

(4) Add $\mathbb{E}(X_1)$ *and* σ^2_X *to* $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *respectively, to obtain* $\mathbb{E}(IV(h(i-1),hi))$ *and* $\mathrm{Var}(IV(h(i-1),hi))$.

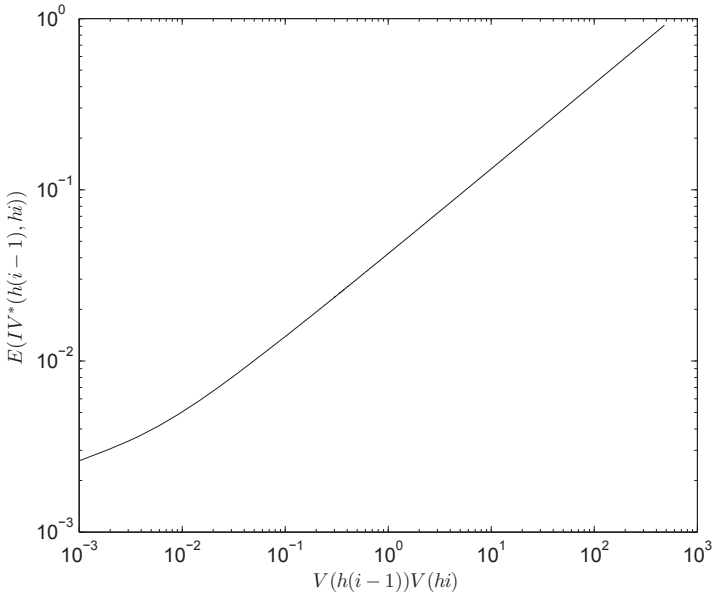

FIGURE 5.1. A log-log graph of $\mathbb{E}(IV^*(h(i-1),hi))$ as a function of $V(h(i-1))V(hi)$.

The use of linear interpolation is justified by the fact that the terms $\mathbb{E}(X_1)$ and $\sigma^2_{X_1}$ have a bigger impact on the moments, according to Tse and Wan (2010). Their argument is also based on the fact that when $\mathbb{E}(IV^*(h(i-1),hi))$ and $\mathrm{Var}(IV^*(h(i-1),hi))$ are regarded as functions of $V(h(i-1))V(hi)$, their graphs

look like piecewise linear curves on a log-log scale. Figures 5.1 and 5.2 illustrate this behavior.

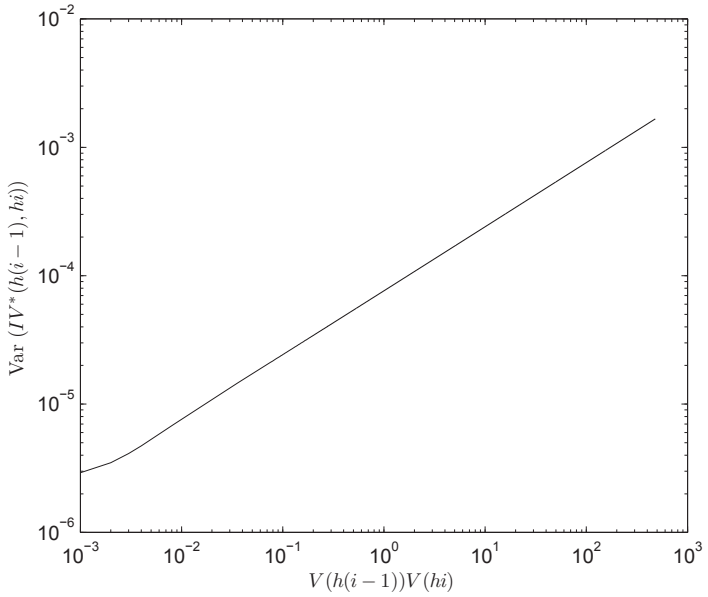

FIGURE 5.2. A log-log graph of $\mathbf{Var}(IV^*(h(i-1),hi))$ as a function of $V(h(i-1))V(hi)$.

5.2.2. Probability density function of $IV(h(i-1),hi)$

To build the p.d.f. of $IV(h(i-1),hi)$'s cache, we first define two vectors. Let $\tilde{v}$ and $\tilde{x}$ be two index vectors. These two vectors are the bases of our cache and they are defined in detail later in this subsection. Figure 5.3 illustrates how we use these vectors. We know from Proposition 2.10.1 that the p.d.f. of $IV(h(i-1),hi)$ depends on $V(h(i-1))$ and $V(hi)$. Thus, we use $\tilde{v}$, our index vector of the variance, as the "height" and the "width" of our cache.

Using Proposition 2.10.1, the p.d.f. is evaluated for fixed values of $V(h(i-1))$ and $V(hi)$ (taken from $\tilde{v}$) and for every element of $\tilde{x}$. Once this is completed, we cache the p.d.f. (note that each dashed line in Figure 5.3 corresponds to a computed p.d.f. on the domain $\tilde{x}$). We now explain the details of our approach.

Definition 5.2.2. *Let $\tilde{v}$ be a vector that contains p monotonic equidistant elements. Its minimum value is defined using the* 0.5 *percentile of the distribution*

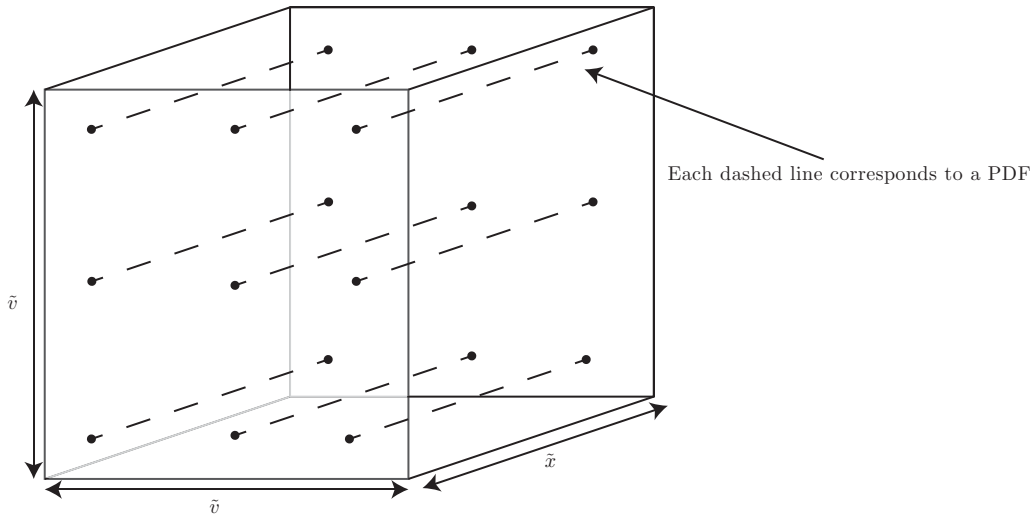

FIGURE 5.3. Illustration of how the cache is built. Note that the "number of p.d.f." (dashed line) in the box is determined by the number p explained in Subsection 5.5.2.

defined in Proposition 2.7.1 at maturity T. Its maximum value is defined using the 99.5 *percentile of the distribution defined in Proposition 2.7.1 at the maturity.*

We also need to define $\tilde{x}$, which is done arbitrarily. In our example, we use a sequence from 0 to 100 using steps of 0.001 such that $\tilde{x} = \{0, 0.001, ..., 100\}$. The maximum seems to be large enough for our applications; however, if there was any reason to suspect that this is not enough, a user could easily increase it up to 1000.

Computing the p.d.f. for every element of $\tilde{x}$ is too intensive; we thus use a trick in order to decrease the computation time needed to build our cache. We compute the value of the p.d.f. at n (see Subsection 5.5.1) points. We then use spline interpolation to recover the non-linear behavior of the p.d.f., for every element of the $\tilde{x}$ vector.

Definition 5.2.3. *We define the n evaluation points as an equidistant monotonic sequence starting at $\mathbb{E}(IV(t,T)) - 8\sqrt{\text{Var}(IV(t,T))}$ (see Proposition 2.10.3) and ending at $\mathbb{E}(IV(t,T)) + 8\sqrt{\text{Var}(IV(t,T))}$.*

The "16 standard errors" choice as the domain is arbitrary; though, a change in this assumption does not seems to have crucial consequences here.

5.3. Construction of the proposal density

Our MCMC-based scheme relies on the basics of Broadie and Kaya (2006). In order to speed up the second step of this algorithm, we use a MCMC algorithm called Metropolis-Hastings (see Algorithm 3.3.1). However, to do so, we need to select an appropriate proposal density. This density (denoted q in Section 3.3) should be a reasonable approximation of the target density in order to have an efficient convergence.

The construction of the proposal density is based, essentially, on observations made by Tse and Wan (2010). According to these authors, the integrated variance over time is asymptotically distributed as a moment-matched inverse Gaussian (see Proposition 4.10.1). Thus, we can use this distribution as a starting point.

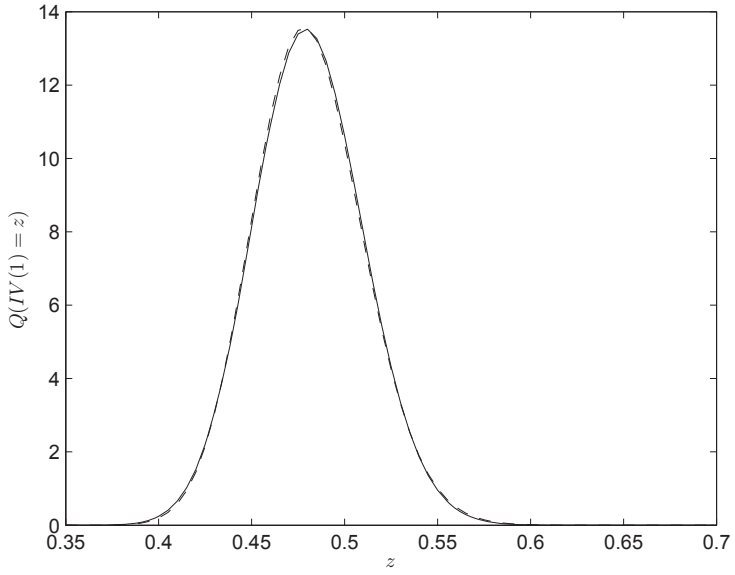

Figure 5.4. A first numerical example: a comparison of the moment-matched inverse Gaussian (dashed line) and the real integrated variance over time p.d.f. (solid line) when the Feller condition is satisfied. We use $\theta = 0.04$, $\kappa = 1$, $\sigma = 0.25$, $T - t = 1$, $V(t) = 0.04$, $V(T) = 0.04$.

When the Feller condition is satisfied, we observe that the behavior of the inverse Gaussian is close to the real p.d.f. (see Figures 5.4 and 5.5 for numerical

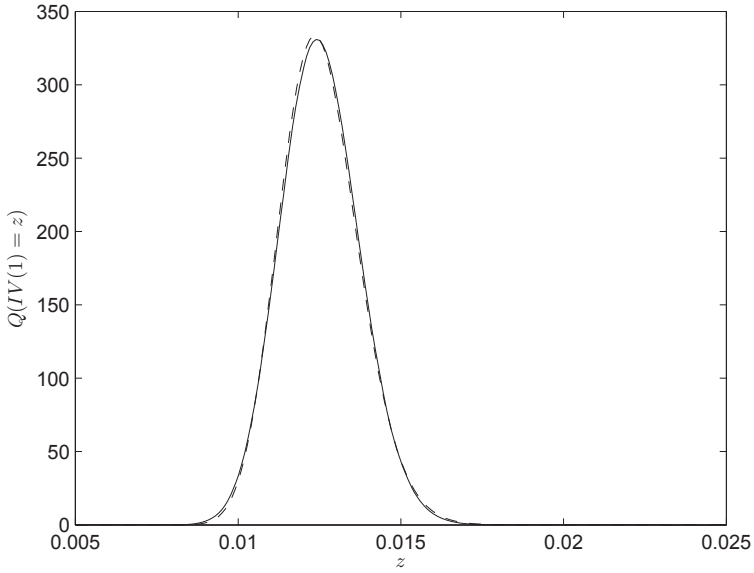

FIGURE 5.5. A second numerical example: a comparison of the moment-matched inverse Gaussian (dashed line) and the real integrated variance over time p.d.f. (solid line) when the Feller condition is satisfied. We use $\theta = 0.04,\ \kappa = 1,\ \sigma = 0.25,\ T-t = 1,$ $V(t) = 0.04,\ V(T) = 0.01.$

examples). Moreover, the behaviors at the mode and in the tails are similar; thus, we shall use this distribution without any modication.

However, when the Feller condition is not satisfied, we observe something different. We notice some disparities between the inverse Gaussian and the real p.d.f. when the Feller condition is violated, when the time step is small, or when the integral bounds are small. A numerical example is presented in Figure 5.6. To improve the proposal density, we propose a mixture of an inverse Gaussian and another density which reproduces the behavior of the real p.d.f. around the origin. The exponential random variable is easy to generate and has a non-zero density at the origin. Thus, it is a good candidate for our mixture.

Definition 5.3.1. *The proposal density q for the Metropolis-Hastings step of our MCMC-based scheme is:*

(1) If the Feller condition is satisfied:

$$X_q \stackrel{d}{=} IG(\mu_{IG}, \lambda_{IG}) \tag{5.3.1}$$

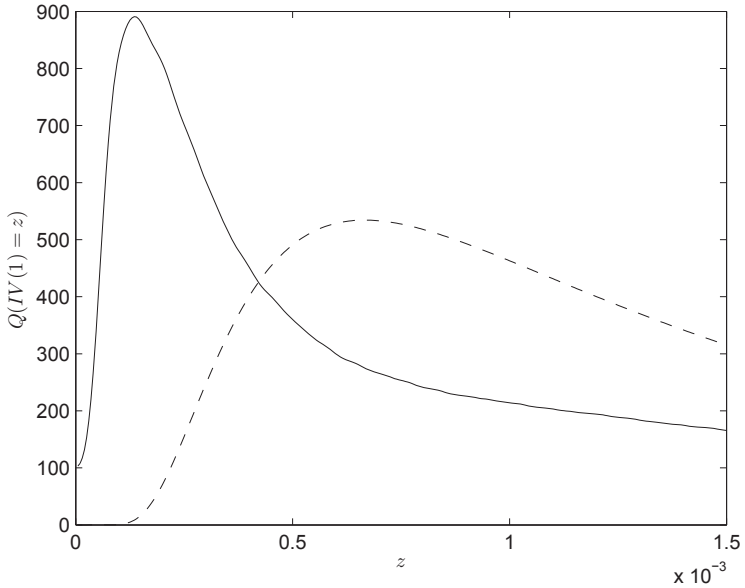

FIGURE 5.6. A comparison of the moment-matched inverse Gaussian (dashed line) and the real integrated variance over time p.d.f. (solid line) when the Feller condition is violated. We use $\theta = 0.04$, $\kappa = 1$, $\sigma = 1$, $T - t = 1$, $V(t) = 0.04$, $V(T) = 0.0004$.

where μ_{IG} *and* λ_{IG} *are moment-matched using the results of Proposition 2.10.3. This proposition yields*

$$\mu_{IG} = \mathbb{E}(IV(h(i-1), hi)) \tag{5.3.2}$$

and

$$\lambda_{IG} = \frac{\mu_{IG}^3}{\mathrm{Var}(IV(h(i-1), hi))}. \tag{5.3.3}$$

(2) If the Feller condition is not satisfied:

$$X_q \stackrel{d}{=} \omega X_{IG}(\mu_{IG}, \lambda_{IG}) + (1-\omega) X_{exp}(\theta_{exp}) \tag{5.3.4}$$

where $\omega \in [0,1]$, X_{IG} *is an inverse Gaussian distribution,* X_{exp} *is an exponential distribution, and* μ_{IG}, λ_{IG} *and* θ_{exp} *are "almost" moment-matched using the results of Proposition 2.10.3. This proposition yields*

$$\mu_{IG} = \mathbb{E}(IV(h(i-1), hi)), \tag{5.3.5}$$

$$\theta_{exp} = \mathbb{E}(IV(h(i-1), hi)) \tag{5.3.6}$$

and

$$\lambda_{IG} = \frac{\mu_{IG}^3}{\text{Var}(IV(h(i-1),hi))}. \tag{5.3.7}$$

The first moment is $\mathbb{E}(IV((h(i-1),hi)))$ since

$$\mathbb{E}(X_q) = \omega\mu_{IG} + (1-\omega)\theta_{exp} = \mathbb{E}(IV((h(i-1),hi))). \tag{5.3.8}$$

However, the second moment is not exactly $\text{Var}(IV((h(i-1),hi)))$ anymore, which does not matter since the focus is on the behavior of the density q. Hence, the variance of q is

$$\text{Var}(X_q) = \mathbb{E}(X_q^2) - \mathbb{E}(X_q)^2 = \omega\,\text{Var}(IV((h(i-1),hi))) + (1-\omega)\mathbb{E}(IV^2((h(i-1),hi))) \tag{5.3.9}$$

which is almost $\text{Var}(IV((h(i-1),hi)))$ if ω is close to 1. In practice, a ω close to 1 is enough to get a good approximation of the behavior of IV.

We have illustrated our proposal density q when the Feller condition is violated to show the difference between ours and the IG density (see Figure 5.7). The proposal density now has a large density around the origin. Thus, it is a better approximation of the real p.d.f. We use $\omega = 0.97$; according to our tests, any value between 0.9 and 1 yields good results.

5.4. Path simulation of the asset price process: a MCMC-based scheme

Our MCMC-based scheme goes as follow.

Algorithm 5.4.1.

(1) Create the caches (see Section 5.2).

(2) $i \leftarrow 1$.

(3) Generate an observation of $\hat{V}(hi)$*, conditionally on* $\hat{V}(h(i-1))$ *using the result of Proposition 2.7.1. This step can be done easily in a numerical computing environment (such as MATLAB).*

(4) Using the Metropolis-Hastings algorithm (see Algorithm 3.3.1), generate a sample value of $\widehat{IV}(h(i-1),hi)$*. The initial element of the chain is the first moment of* $\widehat{IV}((h(i-1),hi)$*, conditional on* $\hat{V}(h(i-1))$ *and* $\hat{V}(hi)$*. In order to speed up the process, we use the cache implemented in Subsection 5.2.2.*

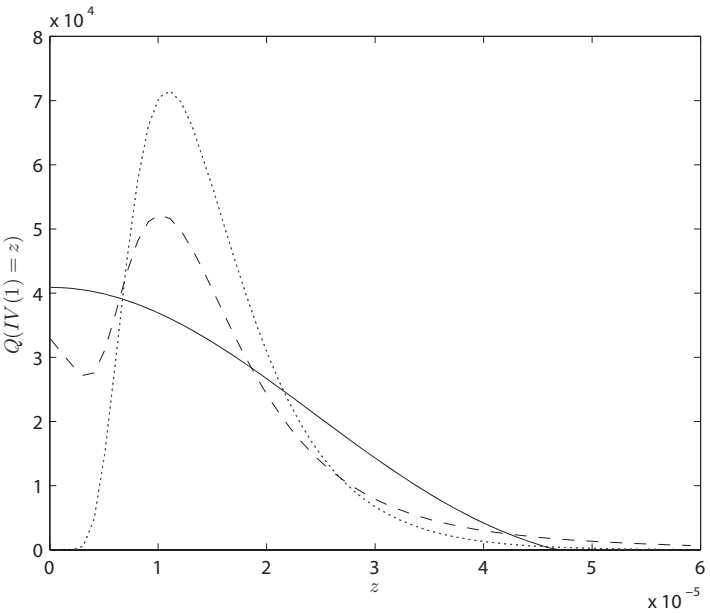

FIGURE 5.7. A comparison of the moment-matched inverse Gaussian (dotted line), the real integrated variance over time p.d.f. (solid line), and the new mixture-based distribution when the Feller condition is violated (dashed line).

The proposal density is described in Section 5.3. Moreover, we detail the optimal "burn-in" period (for our applications) in Subsection 5.5.4.

(5) Generate an independent standardized Gaussian random variable Z_X and set:

$$\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\ dW_X(u) = Z_X \sqrt{\widehat{IV}(h(i-1),hi)}. \tag{5.4.1}$$

(6) Given $\hat{S}(h(i-1))$, $\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\ dW_X(u)$, and $\widehat{IV}(h(i-1),hi)$, compute $\hat{S}(hi)$ using

$$\hat{S}(hi) = \hat{S}(h(i-1)) \exp\left(rh + \frac{\kappa\rho}{\sigma}\widehat{IV}(h(i-1),hi) - \frac{1}{2}\widehat{IV}(h(i-1),hi) \right.$$
$$\left. + \frac{\rho}{\sigma}\Big(\hat{V}(hi) - \hat{V}(h(i-1)) - \kappa\theta h\Big) + \sqrt{1-\rho^2} \int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\ dW_X(u) \right) \tag{5.4.2}$$

(7) Increment i to $i \leftarrow i+1$ and go back to Step 3.

5.5. Simulation scheme tuning

This section presents some adjustments for the MCMC-based scheme. This tuning is obviously not unique. However, according to our tests, it seems to work flawlessly.

5.5.1. Number of evaluation points for the $IV(h(i-1),hi)$'s p.d.f. cache

As we said earlier, in order to speed up the computations, we use few well-chosen evaluation points for the $IV(h(i-1),hi)$'s p.d.f. cache, and then we perform a spline interpolation to recover the p.d.f. evaluated at each element of the vector $\tilde{x}$ (see Subsection 5.2.2). We use a piecewise cubic spline interpolation with n nodes distributed uniformly over the range given in Definition 5.2.3.

We now try to find an appropriate n for our scheme. To do so, we use empirical experiments. Figures 5.8, 5.9, and 5.10 illustrate three examples among hundreds of numerical tests performed, each using different model parameters. When $n = 23$, the spline approximation (dashed line) seems to almost perfectly fit over the real p.d.f. (solid line). If $n < 23$, the fit is not perfect; thus, we shall use $n = 23$ in our scheme since it seems to be the best compromise between computation time and accuracy.

Moreover, we use our tests to compute the mean absolute error (MAE hereafter), defined as

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |f(x_i) - \tilde{f}(x_i)| \tag{5.5.1}$$

where f is the real p.d.f., $\tilde{f}$ is the spline interpolated approximation, and x_i are the elements of $\tilde{x}$. Figure 5.11 illustrates the MAE in function of the number of evaluation points used. As we can see again, $n = 23$ appears to be a good choice.

5.5.2. Size of the $IV(h(i-1),hi)$'s p.d.f. cache

The size of the $IV(h(i-1),hi)$'s p.d.f. cache (called p earlier) is another important factor affecting the computational time for pricing options. We recall that the parameter p introduced in Figure 5.3 is the "width" and the "height" of the cache. Figure 5.12 illustrates the mean absolute error of hundreds of option

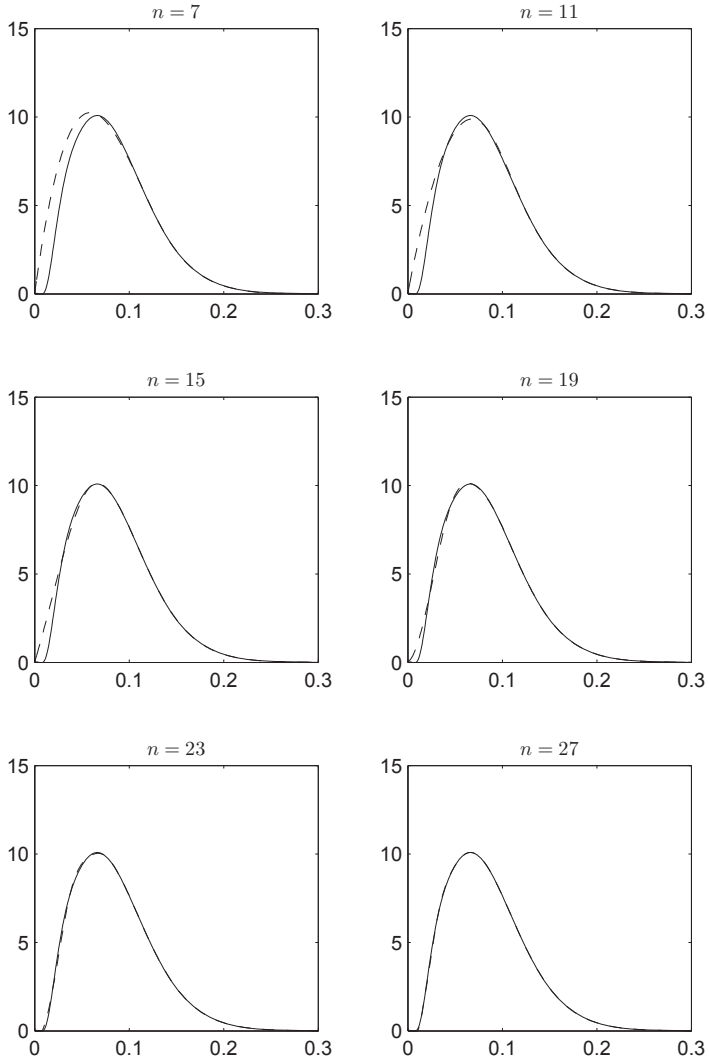

FIGURE 5.8. Approximation using spline (dashed line) and real p.d.f. (solid line) for different n. Here, we use $\theta = 0.04$, $r = 0$, $\kappa = 2$, $\sigma = 1$, $T - t = 5$, $\rho = -0.9$, $V(0) = 0.4$, $V(h) = 0.4$, and $S(0) = 100$.

prices using different p. As we can see, the size of the cache for the p.d.f. of $IV(h(i-1), hi)$ does not seem to have a major impact on the prices. Thus, we advise the reader to use $p \in [5, 15]$ in order to have low numerical biases and low computational times.

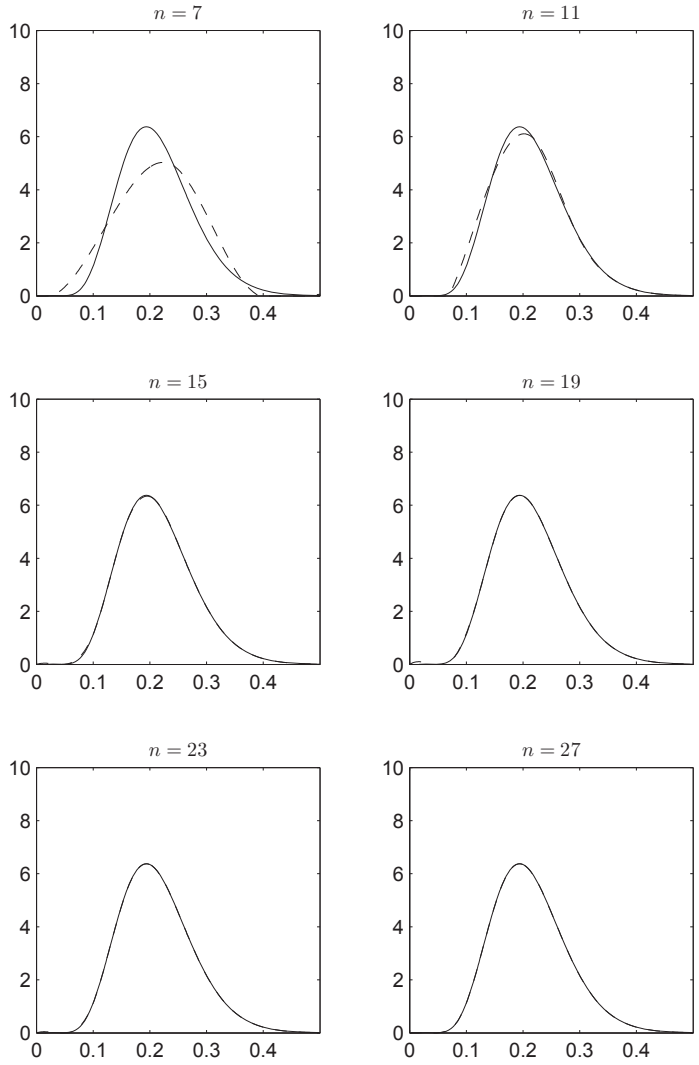

FIGURE 5.9. Approximation using spline (dashed line) and real p.d.f. (solid line) for different n. Here, we use $\theta = 0.04$, $r = 0$, $\kappa = 2$, $\sigma = 1$, $T - t = 5$, $\rho = -0.9$, $V(0) = 0.004$, $V(h) = 0.4$, and $S(0) = 100$.

5.5.3. Size of the $IV(h(i-1), hi)$'s moments caches

The size of the caches for the moments of $IV(h(i-1), hi)$ does not have a major impact on the computation times of our method. Since they are not directly used in our method (we use them as the initial value in the Metropolis-Hastings

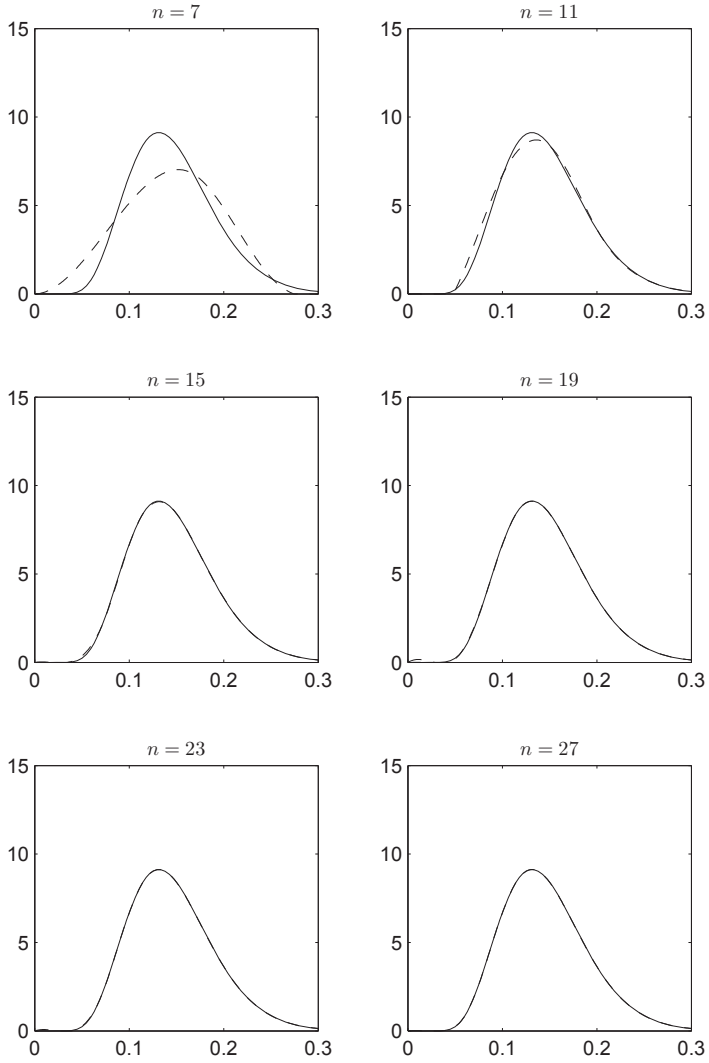

FIGURE 5.10. Approximation using spline (dashed line) and real p.d.f. (solid line) for different n. Here, we use $\theta = 0.04$, $r = 0$, $\kappa = 1$, $\sigma = 0.45$, $T - t = 5$, $\rho = -0.75$, $V(0) = 0.004$, $V(h) = 0.4$, and $S(0) = 100$.

algorithm) and the computation of such moments is quite fast (compared to the p.d.f. computations), their exactness is not overly important. Accordingly, we use a grid starting at 0 and ending at 100, using steps of 0.001.

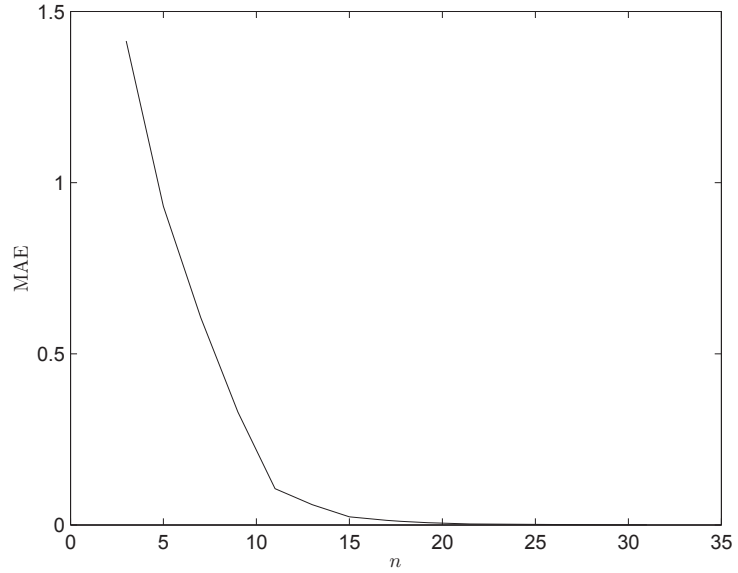

FIGURE 5.11. Average mean absolute error of European option prices as a function of the number of evaluation points used for hundreds of scenarios.

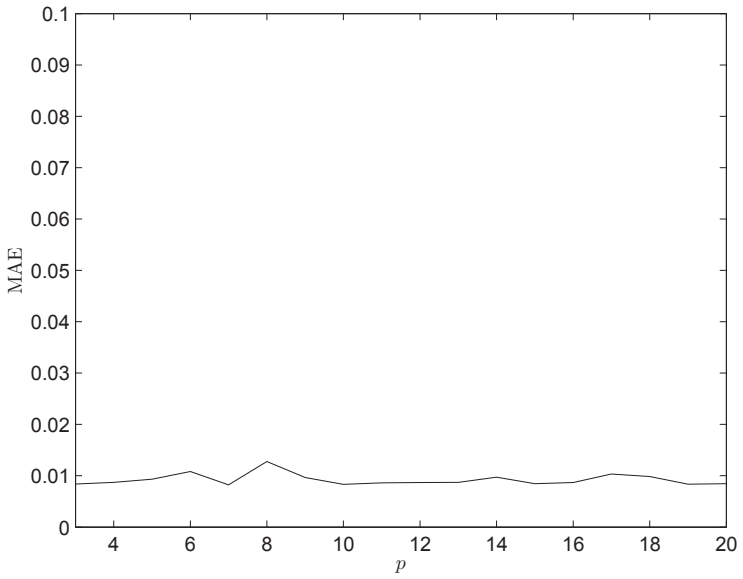

FIGURE 5.12. Average mean absolute error of European option prices as a function of p for hundreds of scenarios.

5.5.4. Number of "burn-in" iterations in the Metropolis-Hastings algorithm

The number of "burn-in" iterations must be at least one since we do not want to keep the first element of the chain (which is arbitrarily chosen). In order to

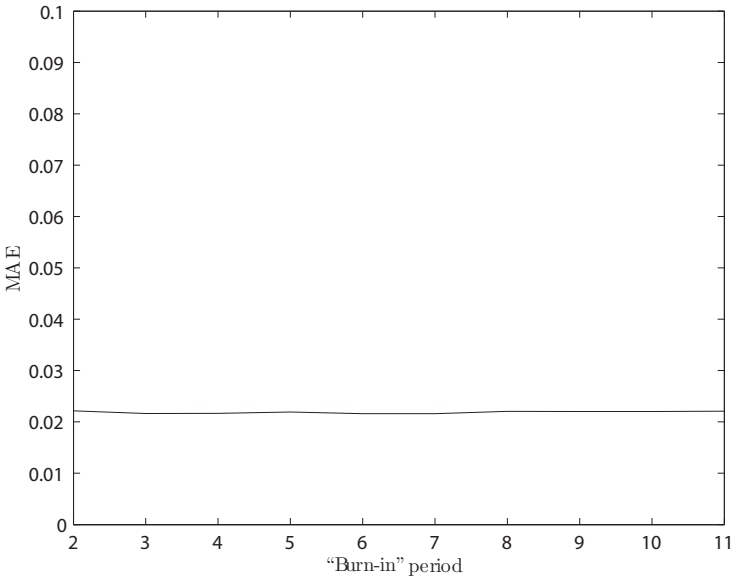

FIGURE 5.13. Average mean absolute error of European option prices as a function of the number of "burn-in" iterations in the Metropolis-Hastings step for hundreds of scenarios.

find the optimal "burn-in" period, we use an empirical approach: we compute the MAE for several scenarios using "burn-in" periods. Figure 5.13 illustrates this empirical test. Thus, we can use any number greater than 1 as the optimal number of "burn-in" iterations; however, 2 seems most efficient since the more "burn-in" iterations we discard, the more time it takes.

It is not surprising that the proposal density q reproduces the real p.d.f.; thus, the candidates from q are almost always accepted (i.e. $\alpha \approx 0.9$ in the Metropolis-Hastings algorithm, see Section 3.3).

5.6. Comparison with popular schemes

In order to evaluate the MCMC-based method described in this chapter, six cases presented in Tables 5.1 and 5.2 are proposed.

The first three cases correspond to those introduced in Andersen (2007). In these cases, the Feller condition (see Proposition 2.7.3) is violated. The main difference between these cases is the strike price. The first one corresponds to an "at the money" option ($K = 100$), the second one is an "out of the money" option ($K = 140$), and the last one is an "in the money" option ($K = 70$).

The next two cases (4 and 5) describe a volatile asset with a strong underlying correlation structure. Case 4 has a ten-year maturity and, Case 5 a five-year maturity. Note that the Feller condition is still violated in these two cases.

The last case was introduced in Smith (2007). In this case, the Feller condition is satisfied.

TABLE 5.1. Heston model parameters Cases 1 to 3.

Case	1	2	3
σ	1	1	1
κ	0.5	0.5	0.5
θ	0.04	0.04	0.04
$V(0)$	0.04	0.04	0.04
ρ	-0.9	-0.9	-0.9
r	0	0	0
T	10	10	10
$S(0)$	100	100	100
K	100	140	70
European call option exact price	13.08467014	0.29577444	35.84976970
Asian call payout option "exact" price	8.19409254	0.02431955	32.62357129

Our running times are obtained by running our MATLAB program on a desktop computer with an overclocked 4.5 GHz Intel Core i7-2600K and 16 GB RAM. We compare our MCMC approach with five algorithms: Andersen (2007)'s quadratic exponential (QE), Milstein scheme (M), Broadie and Kaya (2006) drift interpolation scheme (BK-I), and Zhu (2008)'s transformed volatility (TV) scheme.

TABLE 5.2. Heston model parameters Cases 4 to 6.

Case	4	5	6
σ	1	1	0.5196
κ	1	1	1.0407
θ	0.04	0.04	0.0586
$V(0)$	0.04	0.04	0.0194
ρ	-0.999	-0.999	-0.6747
r	0	0	0
T	10	5	4
$S(0)$	100	100	100
K	100	100	100
European call option exact price	16.70023223	10.98491253	5.58708331
Asian call payout option "exact" price	9.99749579	7.26832512	9.71025320

However, for Asian options, we only compare our scheme to QE and BK-I since TV and M yield poor results. We also note our first MCMC-based scheme "MCMC".

5.6.1. Original MCMC-based scheme, European options

In order to compare our algorithm with the other schemes, we first decided to match computational times. We then ran our scheme first, and then adjusted the number of steps per year, $\frac{m}{T}$, in other schemes so as to yield a comparable computational effort. Tables 5.3, 5.4, 5.5, 5.6, 5.7, and 5.8 provide the biases and computational times for each scheme.

Moreover, the exact prices of our European options are obtained by using Albrecher *et al.* (2007)'s formulation.

Essentially, our scheme seems to compute adequate European call option prices. The results from our MCMC-based scheme are similar to what we get using Andersen's QE and BK-I according to the tables of this subsection. Moreover, our biases are generally lower than QE and BK-I. When MCMC yields larger biases than QE and BK-I, the biases of the three methods are quite similar.

The results obtained using TV and Milstein schemes are poor; large biases are recorded when the Feller condition is violated (Cases 1, 2, 3, 4 and 5).

On Figure 5.14, almost all the cases show convincing results. Each dot represents an observation of a computational time and a relative bias for a given scheme. We compute relative bias and computational time for different sample sizes (from 2^{10} to 2^{19}); the number of steps per year m was determined in order to yield similar results across the schemes. The horizontal lines added are multiple of the standard deviation of the price: the standard deviation of the price is the large-dash line, 3 times the standard deviation is the dashed-dot line, and 5 times the standard deviation is the dotted line. Normally, if we witness a bias less than 3 times the standard deviation, we assume it to be statistically insignificant. Thus, except for Case 2, our scheme produces biases that are statistically insignificant which is a desired property.

So far, we have only considered the option prices. However, we can compare the whole simulated distribution to the real distribution. Figures 5.15, 5.16, 5.17, and 5.18 illustrate how the final stock prices obtained via MCMC, QE, and BK-I are distributed. For all the cases considered here and all the methods used, the simulated c.d.f. seems to be coherent with the analytical c.d.f. of the Heston model. Moreover, we computed the Cramér-Von Mises statistics in order to evaluate the average error committed by our scheme for each case. This statistic is defined by

$$\int_0^\infty \left(\hat{\mathbb{Q}}(S(T) \leq x) - \mathbb{Q}(S(T) \leq x)\right)^2 dx \tag{5.6.1}$$

where $\hat{\mathbb{Q}}(S(T) \leq x)$ is the approximated measure estimated by simulated asset prices and $\mathbb{Q}(S(T) \leq x)$ is the real asset price c.d.f. In order to simplify the

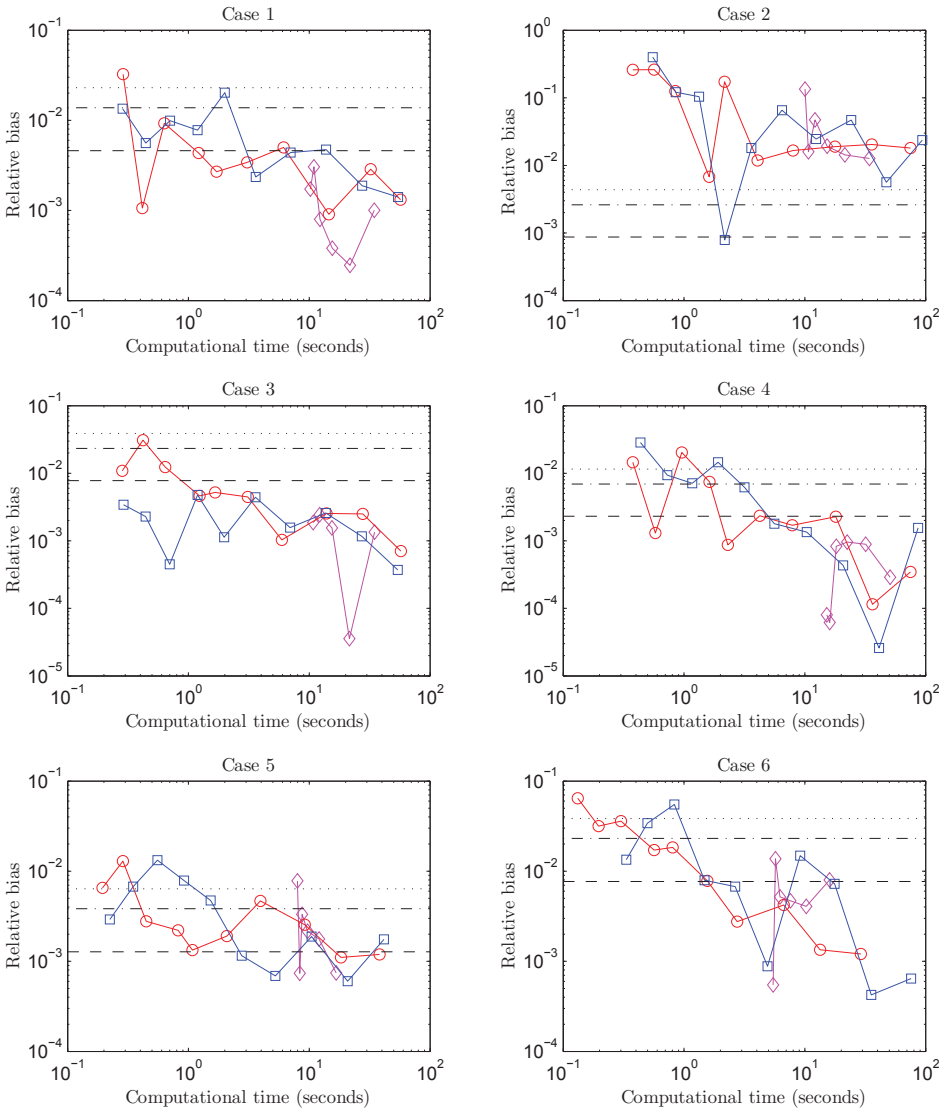

FIGURE 5.14. Relative percentage bias against computational time on pricing European options. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I, and magenta (diamond marker) for MCMC. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line), and 5 times the stardard deviation (dotted line).

TABLE 5.3. Results for the European call option using Case 1.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.202	0.103	0.167	0.078
	Time	10.114	11.079	15.741	34.880
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	1.654	0.284	0.567	0.071
	Time	10.199	11.519	15.206	37.993
M	$\frac{m}{T}$	1500	500	150	100
	Bias (%)	86.713	95.699	108.241	113.187
	Time	10.733	11.474	14.317	39.287
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	0.820	0.566	0.667	0.111
	Time	10.932	10.372	16.345	35.416
TV	$\frac{m}{T}$	1500	600	200	100
	Bias (%)	85.175	94.427	106.103	115.374
	Time	9.835	10.988	17.379	34.863

computation of this statistic, we define a partition $\{x_i\}_{i=0}^{mw} = \{0, \frac{1}{m}, \frac{2}{m}, ..., \frac{mw}{m}\}$ where $\hat{\mathbb{Q}}(S(T) \leq x) = \mathbb{Q}(\hat{S}(T) \leq x) = 0$ for all $x \geq w$. Then,

$$\int_0^\infty \left(\hat{\mathbb{Q}}(S(T) \leq x) - \mathbb{Q}(S(T) \leq x)\right)^2 dx \approx \frac{1}{m} \sum_{i=0}^{mw} \left(\hat{\mathbb{Q}}(S(T) \leq x) - \mathbb{Q}(S(T) \leq x_i)\right)^2. \tag{5.6.2}$$

The Cramér-Von Mises statistics for the MCMC, QE, and BK-I schemes are provided in Table 5.9. This statistic is a function of the x_i's. According to these results, our scheme yields adequate Cramér-Von Mises statistics (in comparison to the other schemes). On average, Cramér-Von Mises statistics are quite similar for each method. This is desirable since, if the distribution of the simulated price is good, all the moments and probabilities computed with this sample should be suitable.

TABLE 5.4. Results for the European call option using Case 2.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	3.455	0.730	0.975	0.497
	Time	10.3121	11.316	15.741	35.291
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	2.604	10.62	5.895	0.245
	Time	10.535	11.260	14.750	38.694
M	$\frac{m}{T}$	1500	500	150	100
	Bias (%)	2993.058	3109.054	3759.531	768.152
	Time	10.542	10.068	14.327	38.915
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	23.572	3.516	3.303	1.692
	Time	11.354	10.472	16.389	34.518
TV	$\frac{m}{T}$	1500	600	200	100
	Bias (%)	2710.549	3006.614	3712.259	4256.525
	Time	10.076	10.882	17.504	35.351

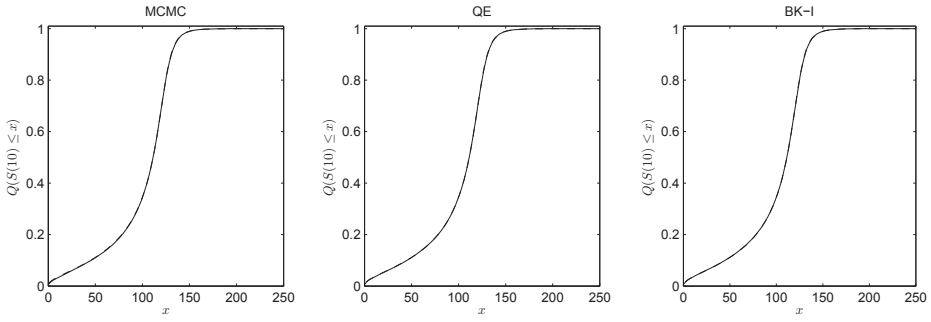

FIGURE 5.15. Comparison of the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Cases 1, 2 and 3.

TABLE 5.5. Results for the European call option using Case 3.

Case 3					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.330	0.108	0.064	0.101
	Time	10.409	11.340	15.808	35.501
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	0.624	0.699	0.307	0.076
	Time	10.719	11.618	15.557	37.514
M	$\frac{m}{T}$	1500	500	150	100
	Bias (%)	21.668	19.694	23.597	24.834
	Time	11.903	9.744	14.018	39.124
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	0.950	0.441	0.282	0.152
	Time	11.381	10.467	16.457	34.738
TV	$\frac{m}{T}$	1500	600	200	100
	Bias (%)	19.387	21.817	22.694	25.475
	Time	10.436	11.099	17.466	34.970

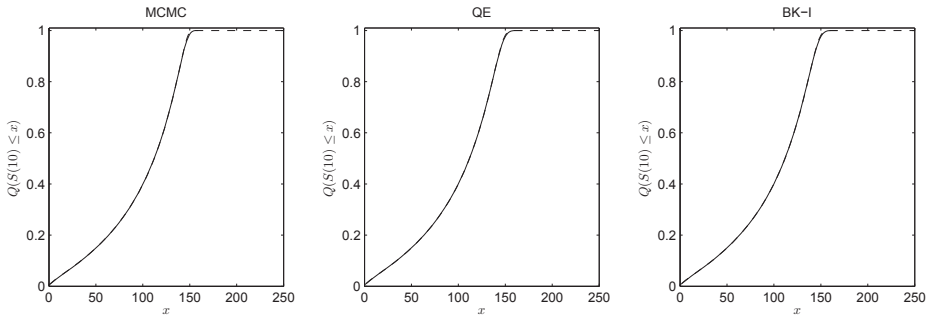

FIGURE 5.16. Comparison of the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Case 4.

TABLE 5.6. Results for the European call option using Case 4.

Case 4					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.904	0.211	0.185	0.068
	Time	14.264	16.257	22.810	51.3433
QE	$\frac{m}{T}$	800	400	150	75
	Bias (%)	0.461	0.258	0.392	0.023
	Time	14.228	18.494	29.067	60.124
M	$\frac{m}{T}$	2000	1000	250	100
	Bias (%)	285.135	311.687	324.124	345.177
	Time	14.124	18.851	21.754	47.147
BK-I	$\frac{m}{T}$	400	200	50	35
	Bias (%)	0.499	0.315	0.378	0.145
	Time	20.136	25.924	21.135	57.004
TV	$\frac{m}{T}$	2000	1000	250	100
	Bias (%)	280.879	298.124	307.145	321.168
	Time	14.348	16.044	22.754	54.145

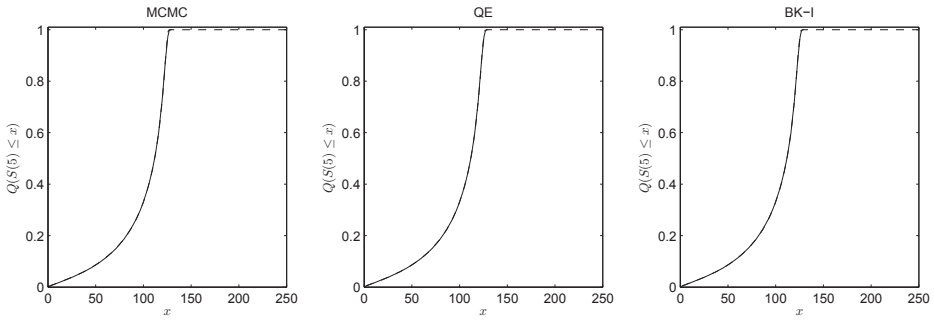

FIGURE 5.17. Comparison of the real c.d.f. (solid line) and the simulated prices $\hat{S}(5)$ (dashed line) for Case 5.

5.6.2. Original MCMC-based scheme, Asian options

Using the cases introduced at the beginning of this section, we now price Asian options (see Section 1.1).

TABLE 5.7. Results for the European call option using Case 5.

Case 5					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.311	0.087	0.0296	0.0354
	Time	8.594	9.214	12.494	23.0603
QE	$\frac{m}{T}$	800	400	150	75
	Bias (%)	0.219	0.205	0.134	0.052
	Time	9.917	9.021	14.576	19.794
M	$\frac{m}{T}$	2000	1000	250	100
	Bias (%)	290.688	297.452	304.25	214.455
	Time	5.838	9.414	12.904	20.442
BK-I	$\frac{m}{T}$	400	200	50	20
	Bias (%)	0.127	0.224	0.112	0.052
	Time	9.997	9.455	12.056	20.009
TV	$\frac{m}{T}$	2000	1000	250	100
	Bias (%)	285.952	293.154	299.441	306.219
	Time	8.897	9.884	12.459	21.102

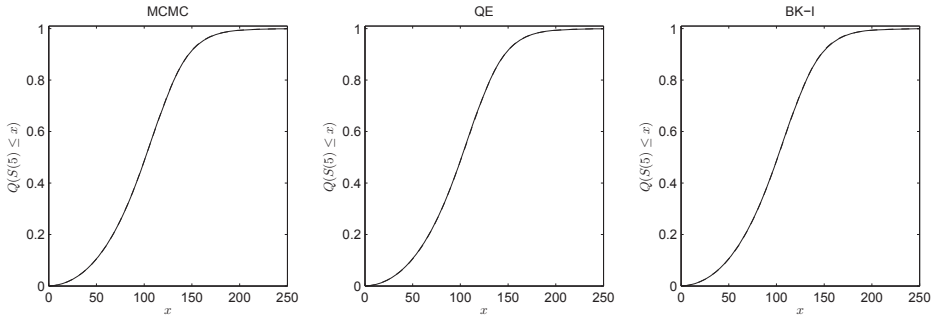

FIGURE 5.18. Comparison of the real c.d.f. (solid line) and the simulated prices $\hat{S}(4)$ (dashed line) for Case 6.

The Asian call payout options, which we refer to as Asian options in this thesis, are defined as follow:

- The average used is the arithmetic mean.

TABLE 5.8. Results for the European call option using Case 6.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.647	0.501	0.414	0.257
	Time	5.636	5.171	7.884	15.124
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	2.416	1.113	0.423	0.104
	Time	4.22	3.98	6.023	13.755
M	$\frac{m}{T}$	1500	500	150	100
	Bias (%)	0.478	0.212	0.241	0.222
	Time	4.768	4.016	6.418	16.340
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	1.892	1.000	0.442	0.312
	Time	5.373	6.353	8.605	17.696
TV	$\frac{m}{T}$	1500	600	200	100
	Bias (%)	0.737	1.142	1.111	0.705
	Time	4.423	4.397	7.450	13.832

TABLE 5.9. Cramér-Von Mises statistics for MCMC, QE and BK-I schemes. We use a sample size of 2^{19} and similar computational times for each method.

Case	MCMC	QE	BK-I
1	2.21×10^{-4}	1.12×10^{-4}	1.67×10^{-4}
2	1.64×10^{-4}	1.12×10^{-4}	1.77×10^{-4}
3	1.11×10^{-4}	1.26×10^{-4}	1.53×10^{-4}
4	2.90×10^{-4}	5.37×10^{-4}	3.97×10^{-4}
5	1.42×10^{-4}	1.65×10^{-4}	0.40×10^{-4}
6	2.32×10^{-4}	0.96×10^{-4}	1.05×10^{-4}

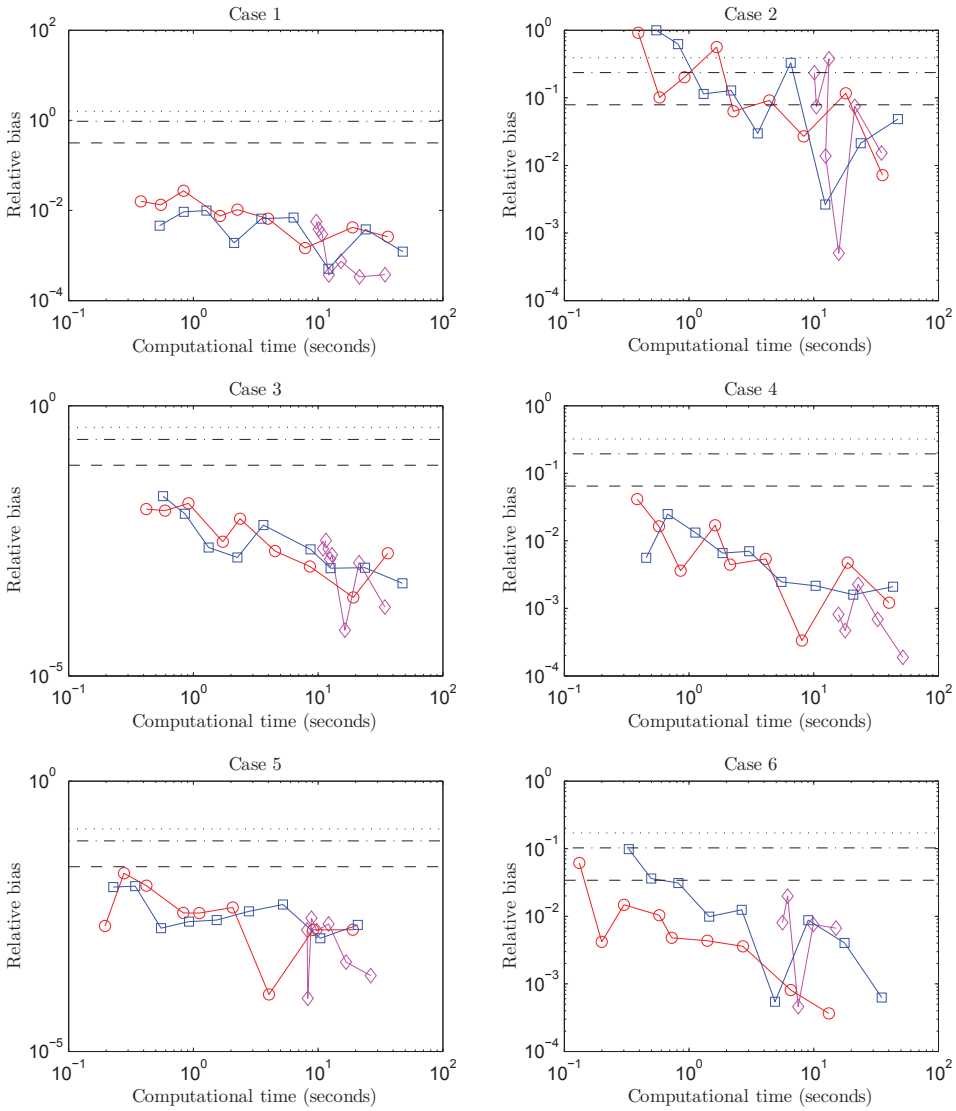

FIGURE 5.19. Relative percentage bias against computational time on pricing Asian options. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I and magenta (diamond marker) for MCMC. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line) and 5 times the stardard deviation (dotted line).

- We take one price per year (at the end). For example, if the maturity of the option is $T \in \mathbb{N}$, then the average, noted $M(T)$, is

$$M(T) = \frac{\sum_{i=1}^{T} \hat{S}(i)}{T}. \tag{5.6.3}$$

TABLE 5.10. Results for the Asian call payout option using Case 1.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	1.004	0.742	0.321	0.036
	Time	10.424	11.229	15.716	36.254
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	0.924	0.376	0.232	0.214
	Time	10.327	11.133	15.472	38.456
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	1.750	0.706	0.197	0.128
	Time	10.912	10.382	16.303	35.642

- The payout at time T is defined by

$$A(K, T-T, \Theta) = \max(M(T) - K, 0) \tag{5.6.4}$$

where K is the strike price and A is the price of the Asian option. Thus, at time 0, it is

$$A(K, T, \Theta) = e^{-rT}\mathbb{E}^{\mathbb{Q}}(\max(M(T) - K, 0)). \tag{5.6.5}$$

Thus, using this definition and our scheme, we are now able to compute prices for this contingent claim. Tables 5.10, 5.11, 5.12, 5.13, 5.14, and 5.15 show the biases and the computational times of each scheme.

The "exact" prices of our Asian options are obtained with the BK-I scheme using 2^{25} paths and 24 steps per year.

Essentially, our scheme seems to compute adequate Asian call option prices. The results from our MCMC-based scheme are similar to what we get using Andersen's QE and BK-I according to the tables of this subsection. For Case 2, the percentage relative biases seem a bit higher; however, the real price of this option is 0.02431955; thus, this strange behavior is caused by the division of a very small number.

On Figure 5.19, almost all the cases show convincing results. The majority of the biases are lower than one standard error, thus almost all the bias are statistically insignificant. However, the results of Case 5 could be more convincing.

TABLE 5.11. Results for the Asian call payout option using Case 2.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	34.607	1.726	4.069	2.476
	Time	10.313	11.229	15.916	35.011
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	10.059	10.391	15.657	9.228
	Time	10.416	11.515	15.584	38.437
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	56.551	5.481	9.104	1.275
	Time	10.951	10.333	16.250	35.455

TABLE 5.12. Results for the Asian call payout option using Case 3.

Case 3					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.049	0.117	0.056	0.056
	Time	10.096	10.546	15.741	34.593
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	0.169	0.4249	0.0543	0.070
	Time	10.245	11.126	14.684	40.024
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	0.535	0.153	0.067	0.081
	Time	10.973	10.416	16.444	35.868

TABLE 5.13. Results for the Asian call payout option using Case 4.

Case 4					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.227	0.154	0.133	0.072
	Time	14.566	16.251	23.250	53.149
QE	$\frac{m}{T}$	800	400	150	75
	Bias (%)	1.647	0.118	0.227	0.056
	Time	13.710	18.207	28.749	59.874
BK-I	$\frac{m}{T}$	400	200	50	35
	Bias (%)	0.479	0.667	0.190	0.109
	Time	20.076	25.574	21.22	57.004

TABLE 5.14. Results for the Asian call payout option using Case 5.

Case 5					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.235	0.089	0.044	0.005
	Time	8.331	9.036	12.629	26.857
QE	$\frac{m}{T}$	800	400	150	75
	Bias (%)	0.377	0.117	0.143	0.155
	Time	6.777	9.033	14.481	29.490
BK-I	$\frac{m}{T}$	400	200	50	35
	Bias (%)	0.520	0.680	0.102	0.107
	Time	10.183	12.880	10.591	28.530

TABLE 5.15. Results for the Asian call payout option using Case 6.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC ($p=5$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.874	0.270	0.139	0.204
	Time	5.604	5.959	7.878	16.437
QE	$\frac{m}{T}$	600	250	80	50
	Bias (%)	1.539	0.375	0.122	0.064
	Time	3.601	4.056	5.496	14.109
BK-I	$\frac{m}{T}$	200	75	35	20
	Bias (%)	0.848	0.408	0.699	0.361
	Time	8.484	8.310	11.389	23.496

Chapter 6

MCMC-BASED SIMULATION ALGORITHM BASED ON SMITH'S APPROXIMATION

This chapter discloses the details of a second MCMC-based algorithm. It is based on the approximation elaborated by Smith (2007). First, we discuss how the Smith's (2007) approximation works. Then, the steps of the second scheme are described and the construction of a new cache is explained. After, the tuning for the approximate scheme is given. Finally, the new MCMC-based algorithm is compared with the ones introduced earlier.

6.1. Background information

Instead of using the real p.d.f. found by Broadie and Kaya (2006), we consider Smith (2007)'s approximation. This should speed up the scheme since it decreases the number of variables in the integrated variance's p.d.f. from three to two. Thus, we need to build a two-dimensional cache (instead of a three-dimensional cache).

However, the use of Smith (2007)'s approximation could lead to a significant bias. Figures 6.1 and 6.2 illustrate the real integrated variance over time's p.d.f. (solid line) and the approximated one (dashed line). There is, virtually, no difference when $V(h(i-1))$ and $V(hi)$ are close (or equal). Though, when $V(h(i-1))$ and $V(hi)$ are significantly different, a mismatch between the p.d.f. is observable.

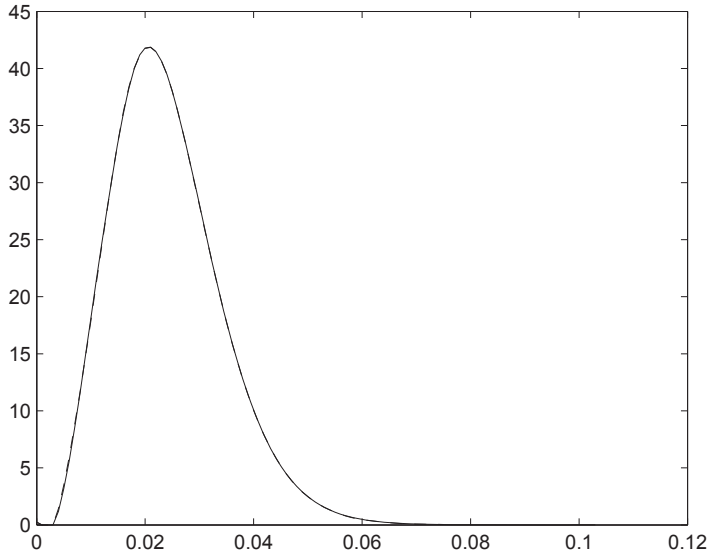

FIGURE 6.1. Comparison of the real integrated variance over time's p.d.f. (solid line) and the approximated one (dashed line) when $V(h(i-1))$ and $V(hi)$ are close to each other. Here, we use $\theta = 0.04$, $\kappa = 1$, $\sigma = 0.45$, $h = 0.5$, $V(h(i-1)) = 0.04$, and $V(hi) = 0.05$.

6.2. Path simulation of the asset price process

We still use the notation introduced in Chapter 5. Our second MCMC-based scheme with Smith's approximation goes as follow.

Algorithm 6.2.1.

(1) Create the $IV(h(i-1),hi)$ p.d.f.'s cache and the moments' caches (see Sections 5.2 and 6.3).

(2) $i \leftarrow 1$.

(3) Generate a sample of $\hat{V}(hi)$, conditionally on $\hat{V}(h(i-1))$, using the result of Proposition 2.7.1. This step could be done easily in a numerical computing environment (such as MATLAB).

(4) Using the Metropolis-Hastings algorithm (see Algorithm 3.3.1), generate a sample of $\widehat{IV}(h(i-1),hi)$. The initial element of the chain is the first moment of $\widehat{IV}((h(i-1),hi)$, conditional on $\hat{V}(h(i-1))$ and $\hat{V}(hi)$. In order to speed up the process, we use the cache implemented in Subsection 6.3.1.

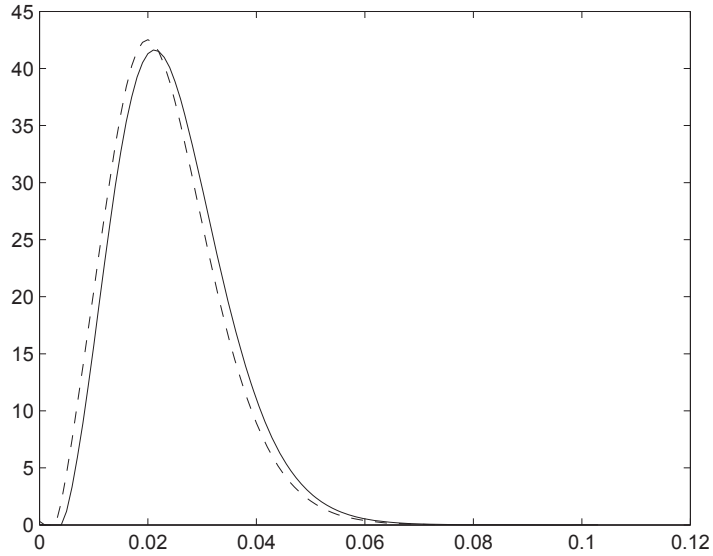

FIGURE 6.2. Comparison of the real integrated variance over time's p.d.f. (solid line) and the approximated one (dashed line) when $V(h(i-1))$ and $V(hi)$ are far from each other. Here, we use $\theta = 0.04$, $\kappa = 1$, $\sigma = 0.45$, $h = 0.5$, $V(h(i-1)) = 0.01$, and $V(hi) = 0.1$.

The proposal density is described in Section 5.3. Moreover, we detail the optimal "burn-in" period (for our applications) in Subsection 6.4.2.

(5) Generate an independent stardardized Gaussian random variable Z_X and set:

$$\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\, dW_X(u) = Z_X \sqrt{\widehat{IV}(h(i-1),hi)}. \tag{6.2.1}$$

(6) Given $\hat{S}(h(i-1))$, $\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\, dW_X(u)$, and $\widehat{IV}(h(i-1),hi)$, compute $\hat{S}(hi)$ using

$$\hat{S}(hi) = \hat{S}(h(i-1)) \exp\left(rh + \frac{\kappa\rho}{\sigma}\widehat{IV}(h(i-1),hi) - \frac{1}{2}\widehat{IV}(h(i-1),hi) \right.$$
$$\left. + \frac{\rho}{\sigma}\left(\hat{V}(hi) - \hat{V}(h(i-1)) - \kappa\theta h\right) + \sqrt{1-\rho^2} \int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\, dW_X(u) \right) \tag{6.2.2}$$

(7) Increment i to $i \leftarrow i+1$ and go back to Step 3.

This scheme is almost identical to the one we introduced in the last chapter.

6.3. Caches implementation

6.3.1. Probability density function of $IV(h(i-1),hi)$

To build this cache, we essentially use the same steps as for the original MCMC-based method. However, here, we only need a two-dimensional cache (instead of a three-dimensional cache). We still use $\tilde{v}$ and $\tilde{x}$, the two index vectors introduced in Chapter 5.

These two vectors are, again, the basis of our cache. Figure 6.3 illustrates how we use these vectors. We use $\tilde{v}$, our index vector for the variance, as the "height" and $\tilde{x}$ as the "depth" of our cache.

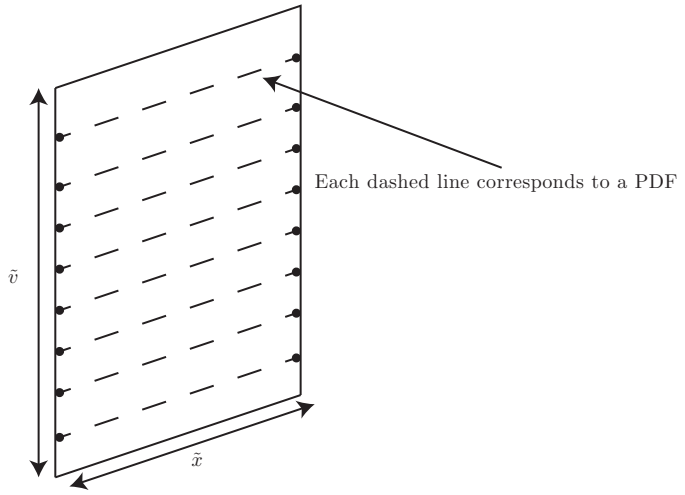

Figure 6.3. Illustration of how the cache is built. Note that the "number of p.d.f." (dashed line) in the box is determined by the number p explained in Subsection 5.5.2.

6.4. Algorithm tuning

We now provide adjustments for our second MCMC-based scheme. These are obviously not unique. Almost all the adjustements provided earlier remain the same. We use the same number of evaluation points for the $IV(h(i-1),hi)$'s p.d.f. cache.

6.4.1. Size of the $IV(h(i-1),hi)$'s p.d.f. cache

The size of the $IV(h(i-1),hi)$'s p.d.f. cache (called p earlier) is another important factor in the time that our method takes to compute the price of an option. We recall that the parameter p introduced in Figure 6.3 is the "height" of the cache.

Figure 6.4 illustrates the mean absolute error of hundreds option prices. As we can see, the size of the $IV(h(i-1),hi)$'s p.d.f. cache does not seems to have a major impact on the prices. Thus, we suggest the reader to use again $p \in [5,15]$ in order to have low numerical biases and low computation times. We will set $p = 10$ in our actual examples.

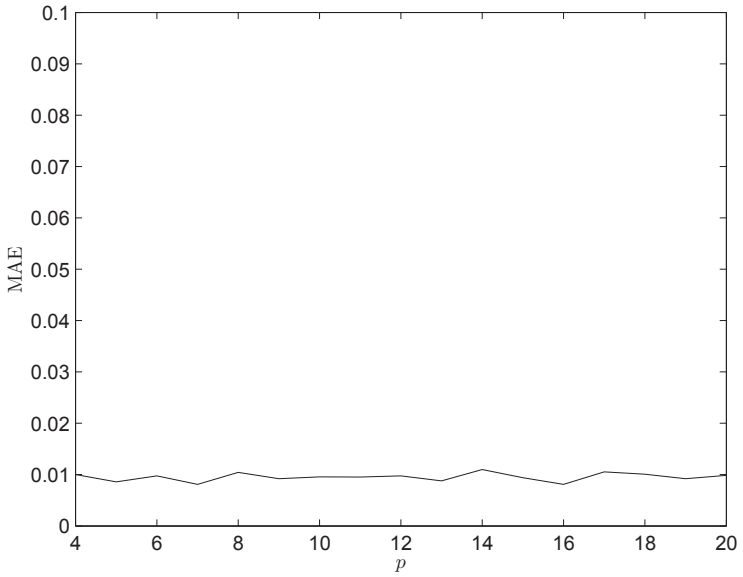

FIGURE 6.4. Mean absolute error of European option prices as a function of p for hundreds of scenarios for the second MCMC-based scheme.

6.4.2. Number of "burn-in" iterations in the Metropolis-Hastings step

Here, again, two iterations seem sufficient. It makes sense: even if we use the approximation, Smith's p.d.f. and the proposal density are close to each

other. Thus, the acceptance probability is around 0.9 for each sample. Figure 6.5 illustrates this fact.

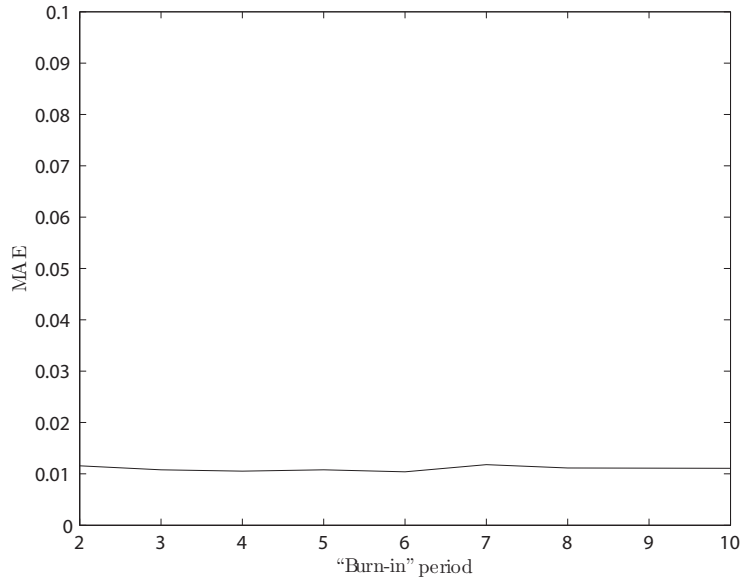

FIGURE 6.5. Mean absolute error of European option prices as a function of the number of "burn-in" iterations in the Metropolis-Hastings step for the second scheme.

6.5. Comparison with popular schemes

In this section, we compare this MCMC-based method to popular algorithms in the literature. In order to evaluate the second MCMC-based method described earlier in this chapter, we use the cases described in Tables 5.1 and 5.2. The approximation scheme using Smith (2007) is noted "MCMC-S".

6.5.1. MCMC-based scheme using Smith's approximation, European options

In order to compare our second scheme with the other ones, we repeat the numerical study performed in Chapter 5; however, we do not consider TV and Milstein schemes since they yield poor results. Tables 6.1, 6.2, 6.3, 6.4, 6.5, and 6.6 show the results of these comparisons.

Essentially, our scheme seems to compute adequate European call option prices. In general, our biases are lower or similar to those found using QE and BK-I, except for Case 6. For this case, the biases found with our scheme are slightly higher than those found by QE and BK-I.

On Figure 6.6, almost all the cases show convincing results; the results from Case 6 (Table 6.6) seem less convincing; however, the biases are, on average, lower than one standard deviation (thus, not significantly different from zero). Once again, the biases in Case 2 (Table 6.2) are five times larger than one standard error. Thus, except for Case 2, all the schemes considered produce biases that are not significantly different from zero, statistically speaking.

Figures 6.7, 6.8, 6.9, and 6.10 illustrate how the final prices obtained via MCMC-S, QE, and BK-I are distributed.

Moreover, we computed, using (5.6.2), the Cramér-Von Mises statistics in order to evaluate the average error committed by our algorithm for each scenario. They are specified in Table 6.7. According to these results, our scheme yields adequate Cramér-Von Mises statistics (in comparison to the other schemes). However, they are slightly higher than those for QE and BK-I.

TABLE 6.1. Results for the European call option using Case 1 and the MCMC-S scheme.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.554	0.137	0.017	0.019
	Time	5.598	6.472	10.047	25.545
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	0.606	0.121	0.243	0.254
	Time	6.035	6.823	9.971	27.327
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	0.988	0.699	0.063	0.104
	Time	5.035	6.669	11.541	25.903

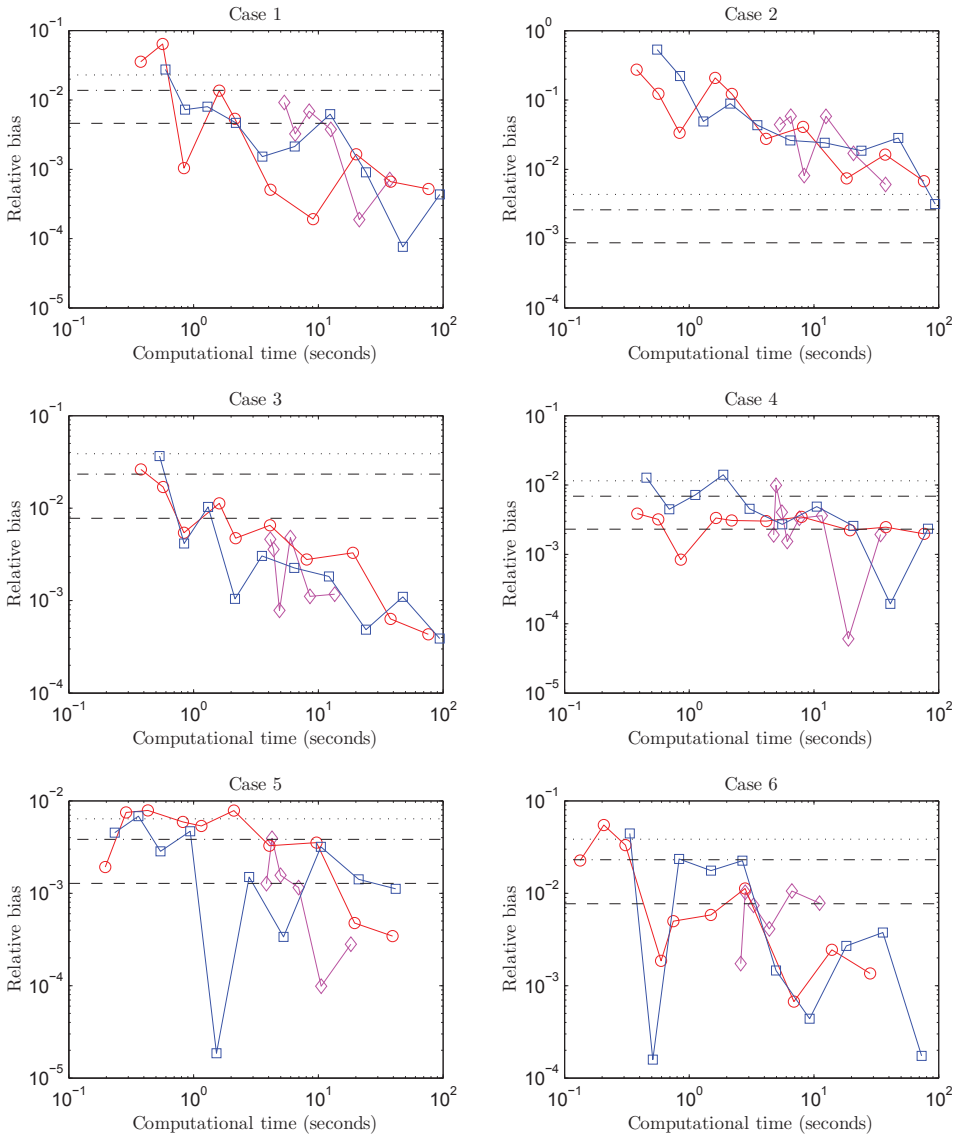

FIGURE 6.6. Relative percentage bias against computational time for pricing European options using MCMC-S scheme. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I, and magenta (diamond marker) for MCMC-S. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line) ,and 5 times the stardard deviation (dotted line).

TABLE 6.2. Results for the European call option using Case 2 and the MCMC-S scheme.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	6.028	0.885	0.206	0.140
	Time	5.558	6.154	9.920	25.084
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	22.512	1.371	1.621	0.788
	Time	5.012	6.719	9.869	28.196
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	1.115	3.088	2.811	1.599
	Time	4.940	6.575	11.390	25.415

TABLE 6.3. Results for the European call option using Case 3 and the MCMC-S scheme.

Case 3					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.465	0.358	0.109	0.133
	Time	5.425	6.225	9.818	24.927
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	0.482	0.195	0.091	0.109
	Time	5.983	6.679	9.789	27.095
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	0.786	0.789	0.407	0.183
	Time	4.933	6.564	11.459	25.386

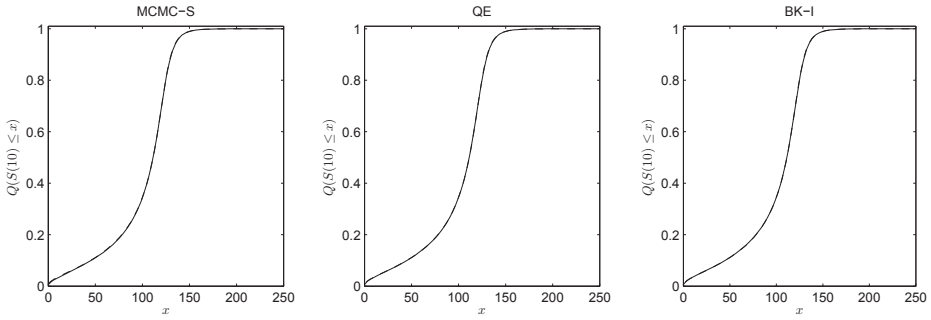

FIGURE 6.7. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Cases 1, 2, and 3 using the MCMC-S scheme.

TABLE 6.4. Results for the European call option using Case 4 and the MCMC-S scheme.

Case 4					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.469	0.400	0.275	0.207
	Time	6.847	7.894	13.377	36.576
QE	$\frac{m}{T}$	450	200	75	50
	Bias (%)	1.445	0.575	0.183	0.139
	Time	7.602	8.758	14.715	39.124
BK-I	$\frac{m}{T}$	150	75	35	20
	Bias (%)	1.139	0.533	0.279	0.151
	Time	6.674	8.713	14.613	32.451

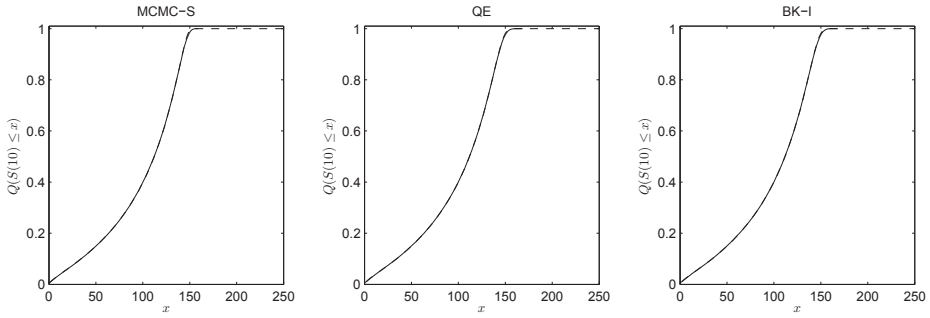

FIGURE 6.8. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Case 4 using the MCMC-S scheme.

TABLE 6.5. Results for the European call option using Case 5 and the MCMC-S scheme.

Case 5					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.042	0.199	0.176	0.158
	Time	5.157	5.726	8.366	19.627
QE	$\frac{m}{T}$	450	200	75	50
	Bias (%)	0.882	0.183	0.502	0.143
	Time	3.812	4.407	7.315	19.433
BK-I	$\frac{m}{T}$	150	75	35	20
	Bias (%)	0.125	0.213	0.131	0.097
	Time	3.468	4.466	7.425	15.866

6.5.2. MCMC-based scheme using Smith's approximation, Asian options

The Asian options we price in this thesis are defined in Subsection 5.6.2. Tables 6.8, 6.9, 6.10, 6.11, 6.12, and 6.13 give biases and computational times for each scheme.

MCMC-S seems to compute adequate Asian call option prices. The results from our MCMC-based scheme are similar to what we get using Andersen's QE

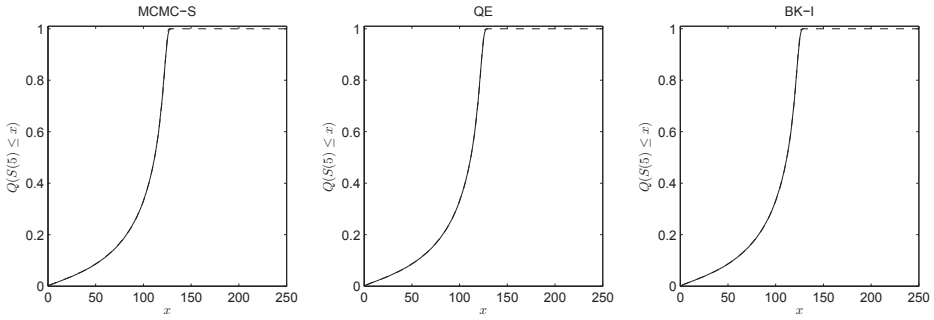

FIGURE 6.9. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(5)$ (dashed line) for Case 5 using the MCMC-S scheme.

TABLE 6.6. Results for the European call option using Case 6 and the MCMC-S scheme.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.169	0.505	0.502	0.620
	Time	3.545	3.991	5.494	12.192
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	2.574	0.181	0.299	0.323
	Time	2.0711	2.411	3.583	10.001
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	2.056	1.408	0.269	0.306
	Time	3.615	5.014	7.724	16.444

and BK-I according to these examples. All the schemes yield, on average, the same biases (for the same computational times).

On Figure 6.11, almost all the cases show convincing results. However, the results of Case 6 could be more convincing, again: our scheme seems to perform less efficiently in this case. However, through Cases 1 to 6, the majority of the biases are lower than one standard error; thus almost all the bias are statistically non significant.

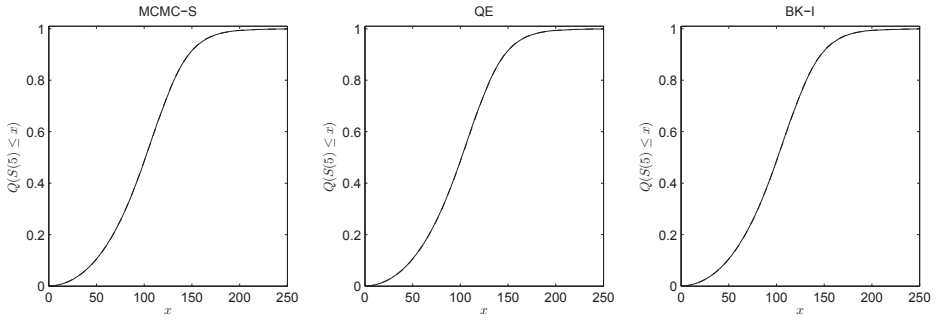

FIGURE 6.10. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(4)$ (dashed line) for the Case 6 using the MCMC-S scheme.

TABLE 6.7. Cramér-Von Mises statistics for MCMC-S, QE, and BK-I schemes. We use a sample size of 2^{19} and similar computational times for each method.

Case	MCMC-S	QE	BK-I
1	3.78×10^{-4}	1.12×10^{-4}	1.67×10^{-4}
2	1.64×10^{-4}	1.12×10^{-4}	1.77×10^{-4}
3	1.56×10^{-4}	1.26×10^{-4}	1.53×10^{-4}
4	6.94×10^{-4}	5.37×10^{-4}	3.97×10^{-4}
5	1.15×10^{-4}	1.65×10^{-4}	0.40×10^{-4}
6	3.54×10^{-4}	0.96×10^{-4}	1.05×10^{-4}

In general, our biases are lower than those found using other methods (for the same computational times). If not, our biases are quite similar to what other methods yield.

As a final note, we could simply say that this method seems more biased than the one implemented in Chapter 5. However, the computation times with

TABLE 6.8. Results for the Asian call payout option using Case 1 and the MCMC-S scheme.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.890	0.126	0.129	0.047
	Time	5.598	6.472	10.047	25.545
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	0.800	0.351	0.119	0.055
	Time	6.035	6.823	9.971	27.327
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	0.988	0.146	0.207	0.040
	Time	5.035	6.669	11.541	25.903

TABLE 6.9. Results for the Asian call payout option using Case 2 and the MCMC-S scheme.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	2.365	1.899	1.304	0.027
	Time	5.558	6.154	9.920	25.084
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	15.326	10.083	1.837	0.484
	Time	5.012	6.719	9.869	28.196
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	40.404	8.266	9.462	0.157
	Time	4.940	6.575	11.390	25.415

the Smith's (2007) method are, in general, smaller than the one obtained at the previous chapter.

TABLE 6.10. Results for the Asian call payout option using Case 3 and the MCMC-S scheme.

MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	0.115	0.179	0.057	0.081
	Time	5.425	6.225	9.818	24.927
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	0.147	0.221	0.074	0.039
	Time	5.983	6.679	9.789	27.095
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	0.110	0.217	0.013	0.103
	Time	4.933	6.564	11.459	25.386

TABLE 6.11. Results for the Asian call payout option using Case 4 and the MCMC-S scheme.

Case 4					
MCMC-S ($p = 10$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.720	0.355	0.265	0.098
	Time	6.847	7.894	13.377	36.576
QE	$\frac{m}{T}$	450	200	75	50
	Bias (%)	0.444	0.254	0.110	0.074
	Time	7.602	8.758	14.715	39.124
BK-I	$\frac{m}{T}$	150	75	35	20
	Bias (%)	0.392	0.098	0.204	0.105
	Time	6.674	8.713	14.613	32.451

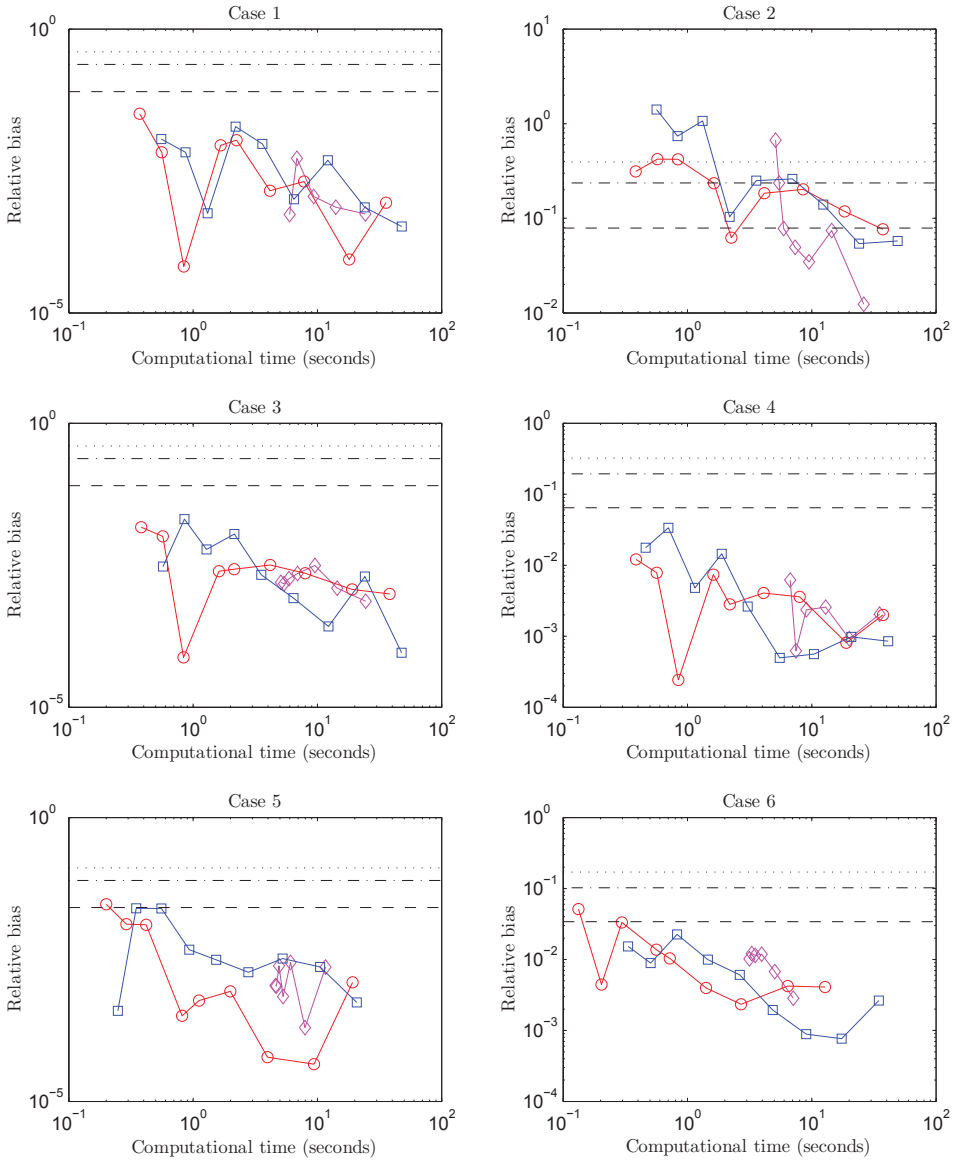

FIGURE 6.11. Relative percentage bias against computational time on pricing Asian options for the second method. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I, and magenta (diamond marker) for MCMC-S. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line), and 5 times the stardard deviation (dotted line).

TABLE 6.12. Results for the Asian call payout option using Case 5 and the MCMC-S scheme.

Case 5					
MCMC-S ($p = 10$)	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.222	0.120	0.015	0.063
	Time	5.157	5.726	8.366	19.627
QE	$\frac{m}{T}$	450	200	75	50
	Bias (%)	0.755	0.092	0.123	0.143
	Time	3.812	4.407	7.315	19.433
BK-I	$\frac{m}{T}$	150	75	35	20
	Bias (%)	1.114	0.301	0.274	0.145
	Time	3.468	4.466	7.425	15.866

TABLE 6.13. Results for the Asian call payout option using Case 6 and the MCMC-S scheme.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
MCMC-S ($p = 10$)	$\frac{m}{T}$	4	4	4	4
	Bias (%)	1.245	0.037	0.277	0.148
	Time	3.545	3.991	5.494	12.192
QE	$\frac{m}{T}$	350	150	50	35
	Bias (%)	1.681	0.845	0.119	0.273
	Time	2.0711	2.411	3.583	10.001
BK-I	$\frac{m}{T}$	100	50	25	15
	Bias (%)	1.054	0.730	0.205	0.254
	Time	3.615	5.014	7.724	16.444

Chapter 7

SIMULATION ALGORITHM BASED ON A GAMMA APPROXIMATION

This chapter discloses the details of a third simulation algorithm for the Heston model. We note that this scheme is not based on any MCMC method. First, we discuss of the basics of our algorithm. Then, we give the formal steps of our new method. Thirdly, the caches we use in the scheme are described. Finally, a numerical comparison is made.

7.1. Background information

We again base our new scheme on the basics of Broadie and Kaya (2006) scheme. The second step of Broadie and Kaya's (the integrated variance over time step) is modified in order to speed up the whole scheme. The first step could also be speeded up. However, this part is not our main concern here.

We recall that Glasserman and Kim (2011) sucessfully propose a new representation of the integrated variance over time. Using the basics from their method and some Laplace transforms, as we mentioned earlier, Tse and Wan (2010) were able to compute the first two moments of the integrated variance over time, IV.

Moreover, Tse and Wan (2010) mentioned that the IV distribution tends toward a moment-matched inverse Gaussian distribution when $h \to \infty$. However, when we price path-dependent securities, it is more appealing to use a small value for h. Thus, the fact that it tends to an inverse Gaussian when $h \to \infty$ is not necessarily convincing in the present case.

From Proposition 2.10.2, γ_n of Equation (2.10.11),

$$\gamma_n = \frac{\kappa^2 h^2 + 4\pi^2 n^2}{2\sigma^2 h^2}, \tag{7.1.1}$$

goes quickly to infinity when h is small, as n increases. As one could see,

$$\gamma_n < \gamma_{n+1} \Leftrightarrow n^2 < (n+1)^2. \tag{7.1.2}$$

Consequently, $\frac{1}{\gamma_n}$ goes quickly to zero. Thus, the mixtures X_1, X_2, and Z_j of Proposition 2.10.2 consist essentially in a sum of gamma random variables (the first gamma random variable of each of the $\eta+2$ mixtures). The exact expression for the convolution of gamma distributions with different scale parameters is quite complicated. According to Stewart *et al.* (2007), it is possible to approximate this complicated distribution using a simple moment-matched gamma distribution. Thus, intuitive reasonning and visual inspection (see Figures 7.1 and 7.2) lead us to believe that a simple gamma distribution could be a good approximation for the integrated variance over time.

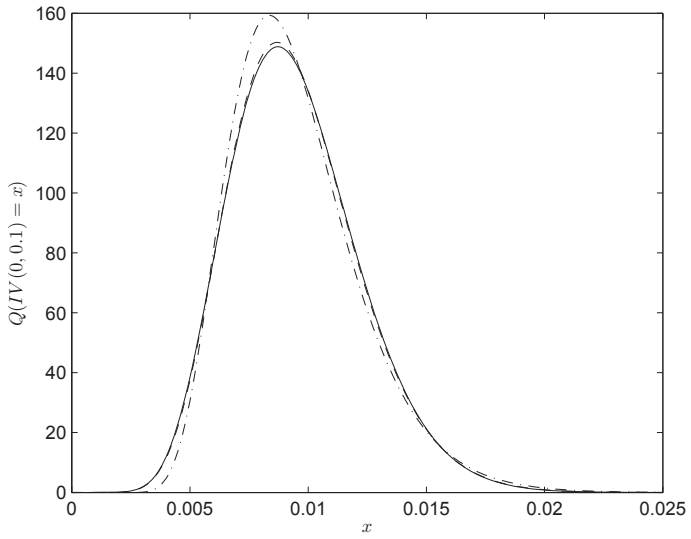

FIGURE 7.1. Comparison between the inverse Gaussian approximation (dashed-dot line), the gamma approximation (dashed line) and the real p.d.f. (solid line) for a small time step (around 0.1)

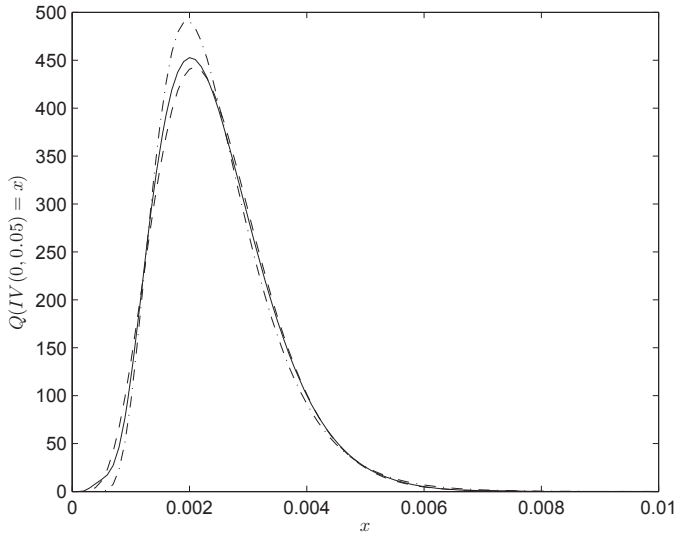

FIGURE 7.2. Comparison between the inverse Gaussian approximation (dashed-dot line), the gamma approximation (dashed line) and the real p.d.f. (solid line) for a small time step (around 0.05).

Our scheme could look like the one implemented by Tse and Wan (2010). Though, we have to ajust the IV's approximated density and the caches, in order to be more efficient.

7.2. Path simulation of the asset price process, third algorithm

We still use the notation introduced in Chapter 5. The third scheme is implemented as follows.

Algorithm 7.2.1.

(1) Create the $IV(h(i-1),hi)$ moments' caches (see Section 7.3 for the new caching method implemented for this scheme).

(2) Initialize $i \leftarrow 1$.

(3) Generate a sample of $\hat{V}(hi)$, conditionally on $\hat{V}(h(i-1))$, using the result of Proposition 2.7.1. This step could be done easily in a numerical computing environment (such as MATLAB).

(4) Given $\hat{V}(hi)$ *and* $\hat{V}(h(i-1))$*, generate a sample of* $\widehat{IV}(h(i-1),hi) \equiv \int_{h(i-1)}^{hi} \hat{V}(u)\,du$ *from a moment-matched gamma distribution using the computed moments available in the caches.*

(5) Generate an independent stardardized Gaussian random variable Z_X *and set:*

$$\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\; dW_X(u) = Z_X \sqrt{\widehat{IV}(h(i-1),hi)}. \tag{7.2.1}$$

(6) Given $\hat{S}(h(i-1))$*,* $\int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\; dW_X(u)$*, and* $\widehat{IV}(h(i-1),hi)$*, compute* $\hat{S}(hi)$ *using*

$$\hat{S}(hi) = \hat{S}(h(i-1)) \exp\left(rh + \frac{\kappa\rho}{\sigma}\widehat{IV}(h(i-1),hi) - \frac{1}{2}\widehat{IV}(h(i-1),hi) \right.$$
$$\left. + \frac{\rho}{\sigma}\Big(\hat{V}(hi) - \hat{V}(h(i-1)) - \kappa\theta h\Big) + \sqrt{1-\rho^2} \int_{h(i-1)}^{hi} \sqrt{\hat{V}(u)}\; dW_X(u) \right) \tag{7.2.2}$$

(7) Increment i *to* $i \leftarrow i+1$ *and go back to Step 3.*

This scheme is almost identical to the ones introduced in the previous chapters, except for the fourth step.

7.3. Caches implementation

We now give a new way to cache the moments of $IV(h(i-1),hi)$. This step is crucial since the distribution we are sampling from depends directly on these moments.

7.3.1. Expected value and variance of $IV(h(i-1),hi)$

We use $\mathbb{E}(IV^*(h(i-1),hi))$ and $\mathrm{Var}(IV^*(h(i-1),hi))$ defined in Definition 5.2.1.

To recover the moments, one could use an approach based on what we implemented previously.

Proposition 7.3.1. *For fast moments calculation, one could use the following steps:*

(1) Precompute $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *using a* ***special grid*** *(see Definition 7.3.1) on the values of* $V(h(i-1))V(hi)$*.*

(2) Compute $\mathbb{E}(X_1)$ *and* $\sigma^2_{X_1}$*.*

(3) Use linear interpolation to approximate $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *from their respective cache.*

(4) Add $\mathbb{E}(X_1)$ *and* σ_X^2 *to* $\mathbb{E}(IV^*(h(i-1),hi))$ *and* $\mathrm{Var}(IV^*(h(i-1),hi))$ *respectively, to obtain* $\mathbb{E}(IV(h(i-1),hi))$ *and* $\mathrm{Var}(IV(h(i-1),hi))$.

The unique difference between this proposition and the one used previously in Chapters 5 and 6 is the grid.

Definition 7.3.1. *The new special grid on the values of* $V(h(i-1))V(hi)$ *is given by*

$$\left\{0, c^{-b}, c^{-b+1}, c^{-b+2}, ..., c^{e-1}, c^{e}\right\} \tag{7.3.1}$$

where c *is a constant choosen between 1 and 2,* b *is the minimal power of* c*, and* e *is the maximum power of* c.

Using this "exponentially spaced" grid, we are able to cache the true behavior of $\mathbb{E}(IV^*(h(i-1),hi))$ and $\mathrm{Var}(IV^*(h(i-1),hi))$. As we said earlier, $\mathbb{E}(IV^*(h(i-1),hi))$ and $\mathrm{Var}(IV^*(h(i-1),hi))$ as a function of $V(h(i-1))V(hi)$ have an almost linear behavior on a log-log scale. Thus, their true behavior is exponential.

Moreover, one could argue to use a spline interpolation to recover the true exponential behavior. Though, it takes too much time and it is an overkill. A simple linear interpolation outperforms the spline interpolation in this case.

Thus, this grid, coupled with linear interpolation, gives good approximations of the moments.

7.4. Comparison with popular schemes

In this section, we compare our third scheme to popular algorithms in the literature. In order to evaluate our third method, we use the cases described in Tables 5.1 and 5.2. The scheme is noted "GA" for gamma approximation.

7.4.1. Gamma approximation scheme, European options

In order to compare our third scheme with the other ones, we repeat the numerical study performed in Chapter 5; however, we do not consider TV and Milstein schemes since they yield poor results. Tables 7.1, 7.2, 7.3, 7.4, 7.5, and 7.6 show the results of these comparisons.

Essentially, our scheme seems to compute adequate European call option prices. In general, our biases are lower than those found using QE and BK-I.

On Figure 7.3, all the cases show convincing results. The biases are, on average, lower than a standard deviation (thus, not significantly different from zero). Once again, the biases in the Case 2 are five times larger than one standard error. Thus, except for Case 2, all the schemes considered produce biases that are not significantly different from zero, statistically speaking.

Figures 7.4, 7.5, 7.6, and 7.7 illustrate how the final prices obtained via GA, QE, and BK-I are distributed. In general, all the graphs in each figure are similar.

Moreover, we computed, using (5.6.2), the Cramér-Von Mises statistics in order to evaluate the average error committed by our algorithm for each scenario. They are specified in Table 7.7. According to these results, our scheme yields adequate statistics (in comparison to the other schemes). Moreover, the results aquired using GA are better (or equivalent) to the ones given by GE and BK-I.

TABLE 7.1. Results for the European call option using Case 1 and the GA scheme.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.723	0.184	0.210	0.051
	Time	0.489	1.384	5.475	21.012
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	2.185	0.361	0.250	0.192
	Time	0.421	1.124	4.878	19.129
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	1.894	0.252	0.345	0.155
	Time	0.482	1.354	5.061	20.118

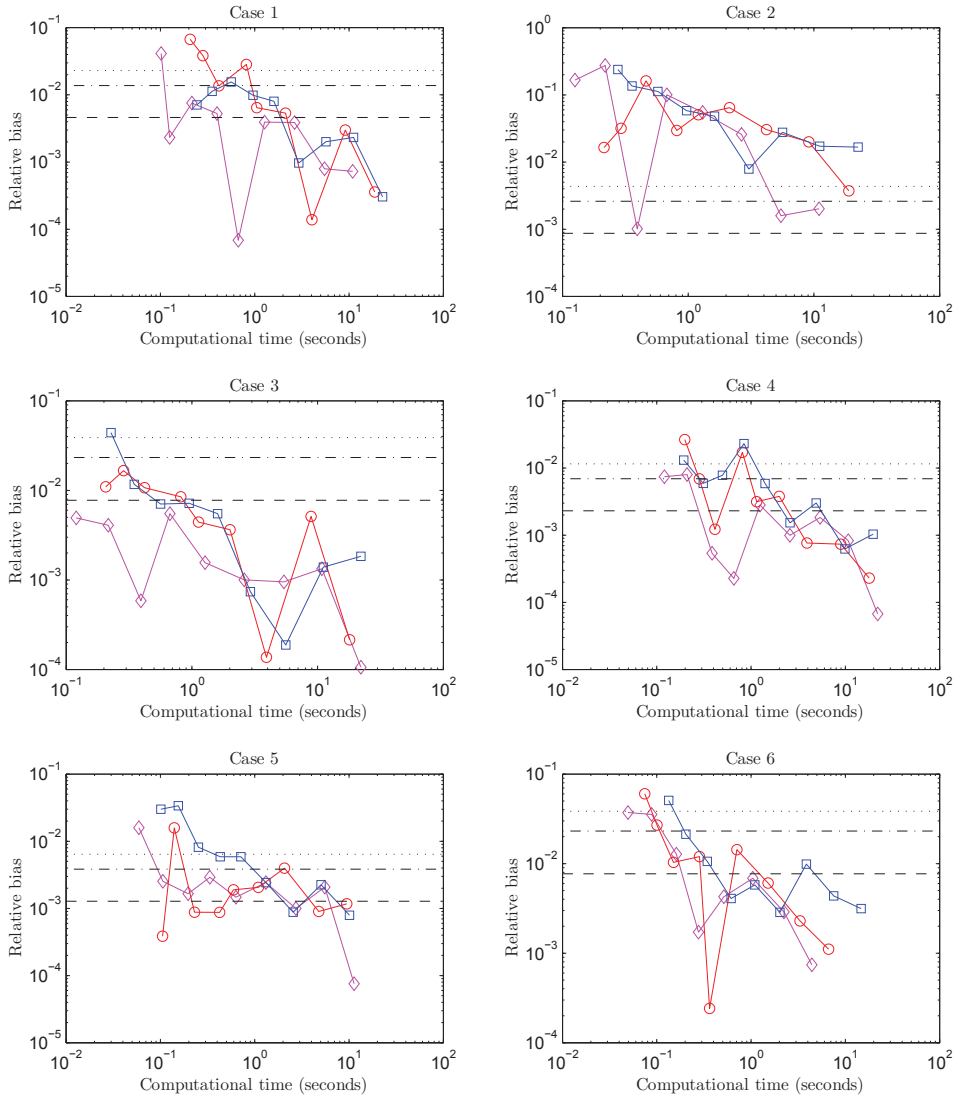

FIGURE 7.3. Relative percentage bias against computational time for pricing European options using GA scheme. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I, and magenta (diamond marker) for GA. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line) ,and 5 times the stardard deviation (dotted line).

TABLE 7.2. Results for the European call option using Case 2 and the GA scheme.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	1.100	2.701	1.746	0.421
	Time	0.487	1.362	5.489	22.0652
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	7.925	2.022	3.541	3.543
	Time	0.463	1.212	4.530	23.454
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	10.911	10.133	3.908	0.677
	Time	0.455	1.335	5.012	25.415

TABLE 7.3. Results for the European call option using Case 3 and the GA scheme.

Case 3					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.240	0.106	0.149	0.064
	Time	0.493	1.362	5.470	22.188
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.848	0.112	0.331	0.105
	Time	0.414	1.561	5.508	22.945
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.559	0.072	0.102	0.060
	Time	0.459	1.321	5.017	21.741

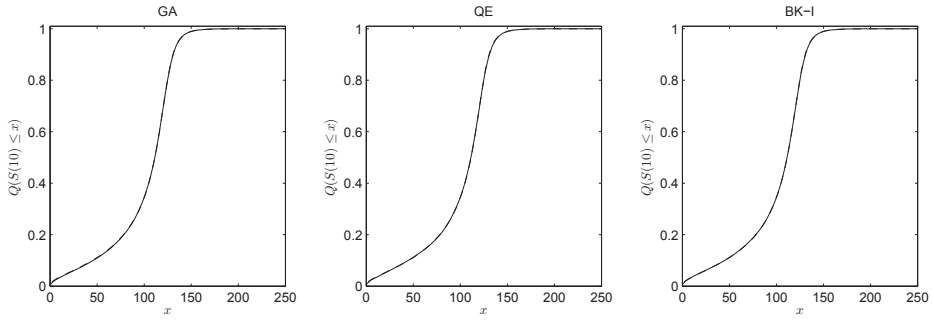

FIGURE 7.4. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Cases 1, 2, and 3 using the GA scheme.

TABLE 7.4. Results for the European call option using Case 4 and the GA scheme.

Case 4					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.738	0.237	0.082	0.041
	Time	0.481	1.365	5.501	22.377
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	1.440	0.951	0.302	0.047
	Time	0.428	1.245	5.717	22.356
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.916	0.416	0.107	0.162
	Time	0.472	1.403	5.248	22.786

7.4.2. Gamma approximation scheme, Asian options

We define, in Subsection 5.6.2, the Asian options we price. Tables 7.8, 7.9, 7.10, 7.11, 7.12, and 7.13 give biases and computational times for each scheme.

GA seems to compute adequate Asian call option prices. The results from our third scheme are similar to what we get using Andersen's QE and BK-I according

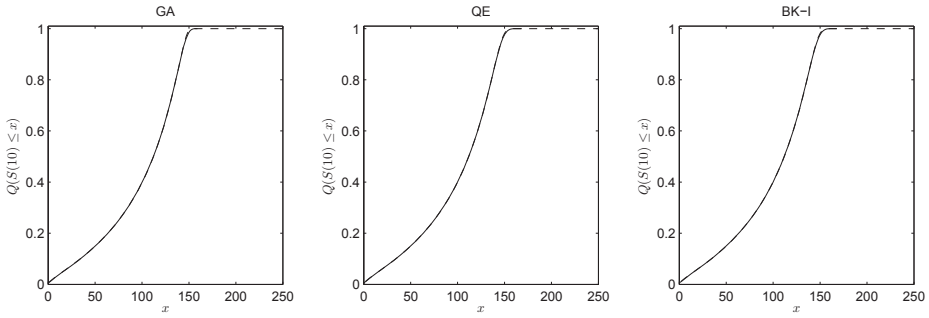

FIGURE 7.5. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(10)$ (dashed line) for Case 4 using the GA scheme.

TABLE 7.5. Results for the European call option using Case 5 and the GA scheme.

Case 5					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.110	0.087	0.021	0.058
	Time	0.282	0.728	2.801	11.187
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.639	0.148	0.220	0.031
	Time	0.217	0.639	2.510	10.652
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	1.683	0.571	0.343	0.043
	Time	0.247	0.719	2.630	10.451

to the tables of this subsection. Moreover, the results using GA seem better in almost all the cases (for comparable computational times).

On Figure 7.8, almost all the cases show convincing results. The results of Case 2 could be more convincing, again. However, through Cases 1 to 6, the majority of the biases are lower than one standard error; thus almost all the biases are statistically non significant.

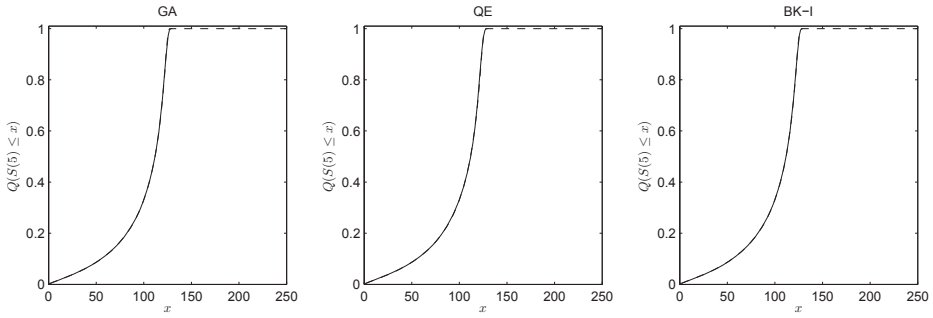

FIGURE 7.6. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(5)$ (dashed line) for Case 5 using the GA scheme.

TABLE 7.6. Results for the European call option using Case 6 and the GA scheme.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.589	0.405	0.344	0.239
	Time	0.244	0.612	2.267	9.221
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.800	0.456	0.210	0.190
	Time	0.265	0.583	2.712	9.115
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.371	0.212	0.123	0.084
	Time	0.251	0.716	2.635	10.470

As a concluding remark, we could point out that this scheme seems better (in term of biases), faster and more intuitive than the two previous ones. The results obtained using this method are, in general, more similar to what the other authors have got in their researches.

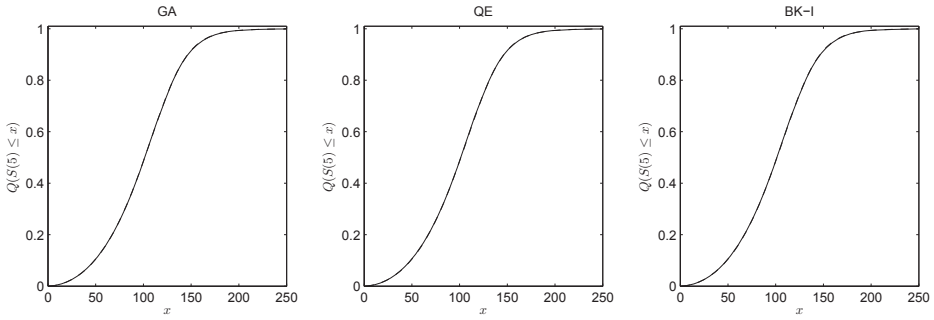

FIGURE 7.7. Comparison between the real c.d.f. (solid line) and the simulated prices $\hat{S}(4)$ (dashed line) for Case 6 using the GA scheme.

TABLE 7.7. Cramér-Von Mises statistics for GA, QE, and BK-I schemes. We use a sample size of 2^{19} and similar computational times for each method.

Case	GA	QE	BK-I
1	1.20×10^{-4}	1.01×10^{-4}	1.28×10^{-4}
2	1.10×10^{-4}	1.12×10^{-4}	1.12×10^{-4}
3	1.10×10^{-4}	1.12×10^{-4}	1.15×10^{-4}
4	7.71×10^{-2}	7.77×10^{-2}	7.71×10^{-2}
5	4.20×10^{-4}	4.27×10^{-4}	4.75×10^{-4}
6	0.96×10^{-4}	1.58×10^{-4}	4.50×10^{-4}

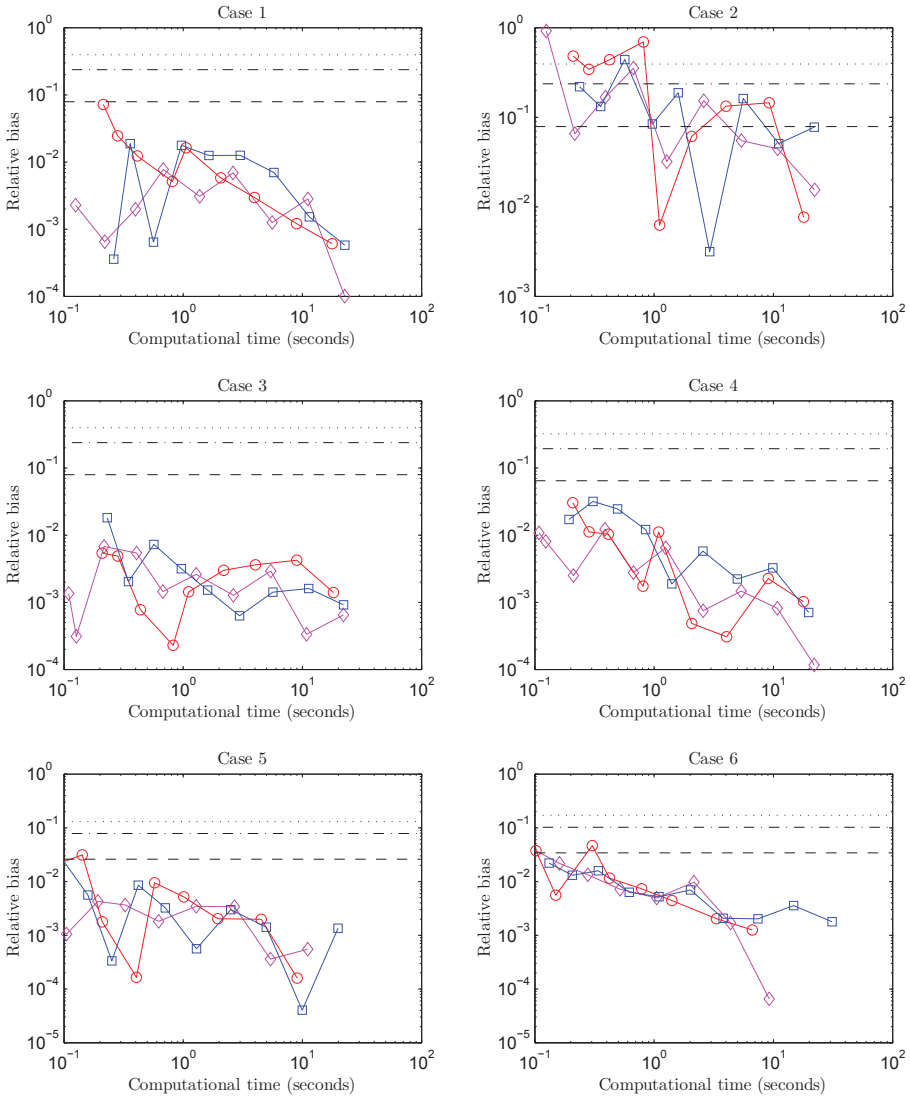

FIGURE 7.8. Relative percentage bias against computational time on pricing Asian options for the second method. The blue line (squared marker) is used for QE scheme, red (circular marker) for BK-I, and magenta (diamond marker) for GA. Moreover, the figure includes the standard deviation of the price (large-dash line), 3 times the standard deviation (dashed-dot line), and 5 times the stardard deviation (dotted line).

TABLE 7.8. Results for the Asian call payout option using Case 1 and the GA scheme.

Case 1					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.666	0.046	0.183	0.029
	Time	0.489	1.384	5.475	21.012
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	1.685	0.407	0.123	0.067
	Time	0.421	1.124	4.878	19.129
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.592	0.457	0.236	0.032
	Time	0.482	1.354	5.061	20.118

TABLE 7.9. Results for the Asian call payout option using Case 2 and the GA scheme.

Case 2					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	1.122	1.033	1.971	1.282
	Time	0.487	1.362	5.489	22.0652
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	4.093	12.500	16.631	2.568
	Time	0.414	1.561	5.508	22.945
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	9.994	3.127	4.768	1.047
	Time	0.455	1.335	5.012	25.415

TABLE 7.10. Results for the Asian call payout option using Case 3 and the GA scheme.

Case 3					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.200	0.232	0.120	0.001
	Time	0.493	1.362	5.470	22.188
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.958	0.224	0.126	0.103
	Time	0.414	1.561	5.508	22.945
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.162	0.602	0.076	0.015
	Time	0.459	1.321	5.017	21.741

TABLE 7.11. Results for the Asian call payout option using Case 4 and the GA scheme.

Case 4					
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.294	0.313	0.065	0.080
	Time	0.481	1.365	5.501	22.377
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.377	0.683	0.725	0.134
	Time	0.428	1.245	5.717	22.356
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	0.896	1.169	0.439	0.083
	Time	0.472	1.403	5.248	22.786

TABLE 7.12. Results for the Asian call payout option using Case 5 and the GA scheme.

Case 5					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.510	0.387	0.137	0.032
	Time	0.282	0.728	2.801	11.187
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	0.985	0.693	0.364	0.179
	Time	0.217	0.639	2.510	10.652
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	2.042	0.409	0.174	0.094
	Time	0.247	0.719	2.630	10.451

TABLE 7.13. Results for the Asian call payout option using Case 6 and the GA scheme.

Case 6					
Method	N	2^{13}	2^{15}	2^{17}	2^{19}
GA	$\frac{m}{T}$	6	6	6	6
	Bias (%)	0.871	0.311	0.034	0.029
	Time	0.244	0.612	2.267	9.221
QE	$\frac{m}{T}$	25	25	25	25
	Bias (%)	1.250	0.521	0.399	0.155
	Time	0.265	0.583	2.712	9.115
BK-I	$\frac{m}{T}$	12	12	12	12
	Bias (%)	1.689	0.765	0.422	0.201
	Time	0.251	0.716	2.635	10.470

CONCLUSION

The main goal of this thesis was to find new ways to simulate stock price paths in the Heston framework. To do so, we described the model in detail. Then, we quoted some important notions about Markov chain Monte Carlo algorithms. After, an exhasutive review on simulation schemes for the Heston model was made. This part helps us understand what are the flaws and the qualities of such schemes. Finally, we introduced three new algorithms: the first two were based on MCMC methods, and the third one was based on a gamma approximation. We gave a thorough analysis of these new algorithms. Moreover, numerical tests were made in order to evaluate the pertinence of MCMC, MCMC-S, and GA.

Our first MCMC algorithm was based on the one implemented by Broadie and Kaya (2006). Instead of using the inversion principle of simulation and the second-order Newton method, we used a Metropolis-Hastings step. To do so, we used the real integrated variance over time p.d.f. given in Broadie and Kaya (2006). The results of this method were generally good, if a little slow.

Our second algorithm was also based on MCMC algorithms. Instead of using the real integrated variance over time p.d.f., we used the approximation made by Smith (2007). This approximation allowed us to reduce the dimension of our p.d.f.'s cache. However, the use of Smith's approximation introduced a small biais in our p.d.f. evaluation. Even if it is more efficient (in a computational way), the Metropolis-Hastings step was still time-consuming.

Finally, our last algorithm was not based on MCMC schemes. Using the basics of Stewart *et al.* (2007), we were able to develop a gamma approximation of the integrated variance over time. Using a moment-matched gamma distribution to sample observations for the integrated variance over time, we were able to find

price paths in an efficient way. This new algorithm is quick and as good as the most popular schemes for the Heston model.

In conclusion, the use of MCMC methods in the simulation of price paths for the Heston model seems potentially justified, but would need to be optimized to really be competitive. The use of an acceptance-rejection algorithm instead of MCMC algorithms could be a clever way to speed up the Metropolis-Hastings step. Some extentions could be made to that effect. Moreover, a deeper analysis of the distribution of the integrated variance over time could lead to another possibility: the simple gamma approximation seems to be adequate, though a more precise density could yield better price paths. The use of gamma mixtures or convolution could give a better fit and could approximate the true behavior of the IV random variable, but whether this could be achieved in a computationally competitive algorithm is still an open question.

BIBLIOGRAPHY

Albrecher, H., Mayer, P., Schoutens, W. and Tistaert, J. (2007). The little Heston trap. *Wilmott Magazine*, 83–92.

Andersen, L. (2007). Efficient simulation of the Heston stochastic volatility model. Technical report, Bank of America.

Bakshi, G., Cao, C. and Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of finance*, 2003–2049.

Baxter, M. and Rennie, A. (1996). *Financial calculus: an introduction to derivative pricing*. London, United Kingdom: Cambridge Univ Pr.

Baz, J. and Chacko, G. (2004). *Financial derivatives: pricing, applications, and mathematics*. London, United Kingdom: Cambridge University Press.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *The Journal of Political Economy*, **81**, 637–654.

Broadie, M. and Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Operations Research*, **54**, 217–231.

Cox, J., Ingersoll Jr, J. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, **53**, 385–408.

Duan, J., Gauthier, G. and Simonato, J. (1999). An analytical approximation for the GARCH option pricing model. *The Journal of Computational Finance*, **2**, 75–116.

Feller, W. (1951). Two singular diffusion problems. *The Annals of Mathematics*, **54**, 173–182.

Gatheral, J. (2006). *The volatility surface: a practitioner's guide*, vol. 357. Wiley.

Gauthier, G. (2010a). Notes du cours calcul stochastique 1: Changement de mesure et théorème de Girsanov. Available at `http://neumann.hec.ca/`

~p240/c8064604/theme_4/12Girsanov.pdf.

Gauthier, G. (2010b). Notes du cours calcul stochastique 1: Espace probabilisé. Available at http://neumann.hec.ca/~p240/c8064604/theme_1/1EspaceProba.pdf.

Gauthier, G. (2010c). Notes du cours calcul stochastique 1: Les processus stochastiques. Available at http://neumann.hec.ca/~p240/c8064604/theme_1/2ProcessusStoch.pdf.

Glasserman, P. and Kim, K. (2011). Gamma expansion of the Heston stochastic volatility model. *Finance and Stochastics*, **15**, 267–296.

Haastrecht, A. and Pelsser, A. (2010). Efficient, almost exact simulation of the Heston stochastic volatility model. *International Journal of Theoretical and Applied Finance (IJTAF)*, **13**, 1–43.

Hakala, J. and Wystup, U. (2002). *Foreign Exchange Risk: models, instruments and strategies*. Risk books.

Harrison, J. and Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, **11**, 215–260.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, **6**, 327–343.

Heston, S. and Nandi, S. (2000). A closed-form GARCH option valuation model. *Review of Financial Studies*, **13**, 585–625.

Hildebrand, F. (1987). *Introduction to numerical analysis*. New York, United States of America: McGraw-Hill.

Hull, J. (1989). *Options, futures, and other derivative securities*. Prentice Hall.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, 281–300.

Hurvich, C. (2000). Some drawbacks of Black-Scholes. Available at http://people.stern.nyu.edu/churvich/Forecasting/Handouts/Scholes.pdf.

Janek, A., Kluge, T., Weron, R. and Wystup, U. (2011). FX smile in the Heston model. In *Statistical Tools for Finance and Insurance*, 133–162. Heidelberg, Germany: Springer.

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

Kotz, S., Johnson, N. and Balakrishnan, N. (2000). *Continuous multivariate distributions: models and applications*, vol. 1. New York, United States of America: Wiley-Interscience.

Lewis, A. (2000). *Option valuation under stochastic volatility*. Newport Beach, United States of America: Finance press, first edn.

Lord, R., Koekkoek, R. and Van Dijk, D. (2008). A comparison of biased simulation schemes for stochastic volatility models. *Tinbergen Institute Discussion Paper No. 06-046/4*.

McKean, H. (1969). *Stochastic integrals*. American Mathematical Society.

Merton, R. (1973). Theory of rational option pricing. *Bell Journal of Economics*, **4**, 141–183.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. *et al.* (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**, 1087–1092.

Michael, J., Schucany, W. and Haas, R. (1976). Generating random variates using transformations with multiple roots. *American Statistician*, **30**, 88–90.

Milstein, G. (1978). A method of second order accuracy integration of stochastic differential equations. *Teoriya Veroyatnostei i ee Primeneniya*, **23**, 414–419.

Moodley, N. (2005). The Heston model: a practical approach with Matlab code. *Faculty of Science, University of the Witwatersrand, South Africa*.

Moro, B. (1995). The full Monte. *Risk*, **8**, 57–58.

Patnaik, P. (1949). The non-central χ^2-and F-distribution and their applications. *Biometrika*, **36**, 202–232.

Revuz, D. (1976). Markov chains. *Bull. Amer. Math. Soc*, **82**, 700–702.

Roberts, G. and Rosenthal, J. (2004). General state space markov chains and mcmc algorithms. *Probability Surveys*, **1**, 20–71.

Rouah, F. (2011). Derivation of the Heston model. Available at `http://www.frouah.com/financenotes/TheHestonModel.pdf`.

Smith, R. (2007). An almost exact simulation method for the Heston model. *Journal of Computational Finance*, **11**, 115–125.

Stein, E. and Stein, J. (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of financial Studies*, **4**, 727–752.

Stewart, T., Strijbosch, L., Moors, H. and Batenburg, P. (2007). A simple approximation to the convolution of gamma distributions. *CentER Discussion Paper No. 2007-70*.

Taleb, N. (2011). *The Black Swan: The Impact of the Highly Improbable*. New York, United States of America: Random House Trade.

Tse, S. and Wan, J. (2010). Low-bias simulation scheme for the Heston model by inverse Gaussian approximation. *SSRN eLibrary*.

Zhu, J. (2008). A simple and exact simulation approach to Heston model. Available at SSRN: http://ssrn.com/abstract.

Appendix A

PROOF OF PROPOSITION 2.1.1

Proposition 2.1.1. *Under the BSM framework, the European call option price at time t is given by*

$$C(S(t),K,T-t,\{r,\sigma\}) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \tag{A.1}$$

where

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}}, \tag{A.2}$$

and $d_2 = d_1 - \sigma\sqrt{T-t}$.

Proof. This demonstration follows the method originally used by Black and Scholes (1973). It is based on deriving and solving the PDE associated to any contingent claim written on the underlying asset. Consider a portfolio with a short position on a call option and α long positions on the underlying asset. The value of this portfolio at time t is given by

$$V(t) = \alpha S(t) - f(t,S(t)) \tag{A.3}$$

where $C(S(t),K,T-t,\{r,\sigma\}) \equiv f(t,S(t)) \equiv f$ is the price of a European call option at time t. Using Itô's Lemma (see Lemma 1.2.1), we have

$$\begin{aligned} dV(t) &= \alpha dS(t) - df(t,S(t)) \\ &= \alpha dS(t) - \left(\frac{\partial f}{\partial S(t)}dS(t) + \frac{\sigma^2}{2}S^2(t)\frac{\partial^2 f}{\partial S^2(t)}dt + \frac{\partial f}{\partial t}dt\right) \\ &= \left(\alpha - \frac{\partial f}{\partial S(t)}\right)dS(t) + \left(\frac{\sigma^2}{2}S^2(t)\frac{\partial^2 f}{\partial S^2(t)} + \frac{\partial f}{\partial t}\right)dt. \end{aligned} \tag{A.4}$$

For the portfolio to be hedged, we must have $\alpha = \frac{\partial f}{\partial S(t)}$. In this case, the portfolio must earn the risk-free rate; thus,

$$dV(t) = rV(t)dt. \tag{A.5}$$

Substituting in the previons equation, we obtain:

$$\frac{\sigma^2}{2}S^2(t)\frac{\partial^2 f}{\partial S^2(t)} + rS(t)\frac{\partial f}{\partial S(t)} + \frac{\partial f}{\partial t} = rf. \tag{A.6}$$

This PDE can theoretically be solved if we specify boundary conditions. In the case of a call option, one must specify the payoff function $f(T,S(T)) = \max(S(T) - K,0)$ (see Section 1.1). This PDE is known as the Fokker-Planck equation. Black and Scholes used a non-intuitive change of variable to solve this PDE. They applied the forward price transformation:

$$\tilde{f}(t,\tilde{S}(t)) = f(t,S(t))e^{r(T-t)} \ , \quad \tilde{S}(t) = S(t)e^{r(T-t)}. \tag{A.7}$$

The different terms in (A.6) become:

$$\frac{\partial f}{\partial t} = \frac{\partial \tilde{f}}{\partial t}e^{-r(T-t)} - rS(t)\frac{\partial \tilde{f}}{\partial \tilde{S}(t)} + r\tilde{f}e^{-r(T-t)}, \tag{A.8}$$

$$\frac{\partial f}{\partial S(t)} = \frac{\partial \tilde{f}}{\partial \tilde{S}(t)} \tag{A.9}$$

and

$$\frac{\partial^2 f}{\partial S^2(t)} = \frac{\partial^2 \tilde{f}}{\partial^2 \tilde{S}^2(t)}e^{r(T-t)}. \tag{A.10}$$

Substituting in (A.6) yields

$$\frac{\partial \tilde{f}}{\partial t} + \frac{1}{2}\sigma^2\tilde{S}^2(t)\frac{\partial^2 \tilde{f}}{\partial \tilde{S}^2(t)} = 0. \tag{A.11}$$

The latter transformation allowed them to eliminate two terms in the original PDE. Thus, the diffusion coefficient still depends on the forward price. We operate now a second transformation as follows:

$$\tilde{S}(t) = K\exp\left(\frac{1}{2}\sigma^2\tau + x\right) \ , \quad \tilde{f}(t,\tilde{S}(t)) = Ky(x,\tau) \tag{A.12}$$

where $\tau = T - t$. The different terms in (A.12) become:

$$\frac{\partial \tilde{f}}{\partial t} = K\frac{\partial y}{\partial t} = K\left(\frac{1}{2}\sigma^2\frac{\partial y}{\partial x} - \frac{\partial y}{\partial \tau}\right) \tag{A.13}$$

$$\frac{\partial \tilde{f}}{\partial \tilde{S}(t)} = K\frac{1}{\tilde{S}}\frac{\partial y}{\partial x} \tag{A.14}$$

and

$$\frac{\partial^2 \tilde{f}}{\partial \tilde{S}^2(t)} = K\frac{1}{\tilde{S}}\frac{\partial y}{\partial x}, \tag{A.15}$$

which gives the following PDE,

$$\frac{\partial y}{\partial \tau} - \frac{1}{2}\sigma^2\frac{\partial^2 y}{\partial x^2} = 0, \tag{A.16}$$

with $y(0,x) \equiv y_0(x) = \max(e^x - 1, 0)$ as the initial condition for y. This new PDE is known as the heat equation in thermodynamic. One can further simplify (A.16) by considering the following change of variable:

$$\tilde{\tau} = \frac{1}{2}\sigma^2\tau, \tag{A.17}$$

yielding the following PDE:

$$\frac{\partial y}{\partial \tilde{\tau}} = \frac{\partial^2 y}{\partial x^2}. \tag{A.18}$$

The fundamental solution of this PDE, or the *Green's function* of diffusion equation, is given by

$$G(\tilde{\tau}, x) = \frac{1}{\sqrt{4\pi\tau}}\exp\left(-\frac{x^2}{4\tau}\right) \tag{A.19}$$

or

$$G(\tau, x) = \frac{1}{\sqrt{2\pi\sigma^2\tau}}\exp\left(-\frac{1}{2}\frac{x^2}{\sigma^2\tau}\right). \tag{A.20}$$

The general solution of (A.18) is written as a convolution product of the *Green's function* and the initial condition:

$$y(\tau, x) = \int_0^\infty G(\tau, x-\omega)(e^x - 1)d\omega. \tag{A.21}$$

For any $\zeta \in \mathbb{R}$, if we define

$$I_\zeta \equiv \int_0^\infty G(\tau, x-\omega)e^{\zeta x}d\omega, \tag{A.22}$$

then $y(\tau, x) = I_1 - I_0$. Thus, we have

$$I_\zeta = \frac{1}{\sqrt{2\pi\sigma^2\tau}}\int_0^\infty \exp\left(-\frac{1}{2}\frac{(x-\omega)^2}{\sigma^2\tau} + \zeta x\right)d\omega = \exp\left(\zeta x + \frac{1}{2}\zeta^2\sigma^2\tau\right)\int_{-\infty}^{d(\zeta)}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\eta^2\right)d\eta \tag{A.23}$$

where

$$\eta = \frac{x - \omega + \zeta\tau\sigma^2}{\sigma\sqrt{\tau}} \tag{A.24}$$

and

$$d(\zeta) = \frac{x + \zeta\tau\sigma^2}{\sigma\sqrt{\tau}}. \tag{A.25}$$

Hence, we obtain

$$I_\zeta = \exp\left(\zeta x + \frac{1}{2}\zeta^2\sigma^2\tau\right)\Phi(d(\zeta)). \tag{A.26}$$

Using the changes of variables backward from y to $\tilde{f}$ and then to f lead to the BSM formula for call prices

$$f(t,S(t)) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \tag{A.27}$$

where

$$d_1 = \frac{\log\left(\frac{S(t)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \tag{A.28}$$

and $d_2 = d_1 - \sigma\sqrt{T-t}$.

□

Appendix B

PROOF OF PROPOSITION 2.4.1

Proposition 2.4.1. *The price of a European call option under the Heston framework is*

$$C(S(t),K,T-t,\{r,\kappa,\theta,V(t),\sigma,\rho\}) = S(t)P_1(\log(S(t)),V(t),T-t) - Ke^{r(T-t)}P_2(\log(S(t)),V(t),T-t), \quad \text{(B.1)}$$

where

$$P_j(\log(S(t)),V(t),T-t) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-\mathrm{i}\phi\log(K)} f_j(\phi,\log(S(t)),V(t))}{\mathrm{i}\phi}\right) d\phi \quad \text{(B.2)}$$

for $j = 1,2$, *and* Re *is the real part of a complex number. The characteristic functions* f_1 *and* f_2 *are given by*

$$f_j(\phi,\log(S(t)),V(t),T-t) = \exp\left(C_j(T-t,\phi) + D_j(T-t,\phi)V(t) + \mathrm{i}\phi\log(S(t))\right), \quad \text{(B.3)}$$

where

$$C_j(T-t,\phi) = ri\phi(T-t) + \frac{a}{\sigma^2}\left((b_j - \rho\sigma\mathrm{i}\phi + h_j)(T-t) - \log\left(\frac{1-g_j e^{h_j(T-t)}}{1-g_j}\right)\right), \quad \text{(B.4)}$$

$$D_j(T-t,\phi) = \frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{\sigma^2}\left(\frac{1-e^{h_j(T-t)}}{1-g_j e^{h_j(T-t)}}\right), \quad \text{(B.5)}$$

$$g_j = \frac{b_j - \rho\sigma \mathrm{i}\phi + h_j}{b_j - \rho\sigma \mathrm{i}\phi - h_j}, \tag{B.6}$$

$$h_j = \sqrt{(\rho\sigma \mathrm{i}\phi - b_j)^2 - \sigma^2(2u_j \mathrm{i}\phi - \phi^2)}, \tag{B.7}$$

$$b_1 = \kappa - \rho\sigma, \quad b_2 = \kappa, \tag{B.8}$$

$$u_1 = \frac{1}{2}, \quad u_2 = -\frac{1}{2}, \tag{B.9}$$

and

$$a = \kappa\theta. \tag{B.10}$$

Proof. This demonstration follows the method of Rouah (2011).

First, we assume that the European call price formula is in the fashion of BSM:

$$\begin{aligned} C(S(t),K,t,T,\Theta) &= e^{-r(T-t)}\, \mathbb{E}^{\mathbb{Q}}(\max(S(T)-K,0)\,|\mathscr{F}_t) && \text{(B.11)} \\ &= e^{-r(T-t)}\, \mathbb{E}^{\mathbb{Q}}(S(T)\mathbb{I}_{S(T)>K}|\mathscr{F}_t) \\ &\quad - e^{-r(T-t)}K\, \mathbb{E}^{\mathbb{Q}}(\mathbb{I}_{S(T)>K}|\mathscr{F}_t) && \text{(B.12)} \\ &= e^{-r(T-t)}P_1 - e^{-r(T-t)}KP_2 && \text{(B.13)} \end{aligned}$$

where $\Theta = \{r,\kappa,\theta,V(t),\sigma,\rho\}$, $X(t) \equiv \log(S(t))$, $P_1 \equiv P_1(X(t),V(t),\tau)$ and $P_2 \equiv P_2(X(t),V(t),\tau)$. In the latter expression, P_2 represents the probability of the call expiring in-the-money, conditional on the value of $X(t)$, $V(t)$ and t. In (B.11), the expected value $\mathbb{E}(\mathbb{I}_{S(T)>K}|\mathscr{F}_t)$ is the probability of the call expiring in-the-money under the risk neutral measure $\mathbb{Q}$. Hence,

$$\mathbb{E}^{\mathbb{Q}}(\mathbb{I}_{S(T)>K}|\mathscr{F}_t) = \mathbb{Q}(S(T)>K|\mathscr{F}_t) = \mathbb{Q}(X(T)>\log(K)|\mathscr{F}_t) = P_2. \tag{B.14}$$

Evaluating $e^{-r(T-t)}\mathbb{E}(S(T)\mathbb{I}_{S(T)>K}|\mathscr{F}_t)$ in (B.11) requires changing the original measure $\mathbb{Q}$ to another measure $\mathbb{Q}^S$. One could define the Radon-Nikodym derivative

$$\mathbb{Z}_t = \frac{S(t)/S(T)}{B(t)/B(T)} = \left.\frac{d\mathbb{Q}}{d\mathbb{Q}^S}\right|_{\mathscr{F}_t} \tag{B.15}$$

where $B(t) = \exp\left(\int_0^t r\ du\right) = e^{rt}$. Then, the second expectation is

$$\begin{aligned} e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}(S(T)\mathbb{I}_{S(T)>K}|\mathscr{F}_t) &= \mathbb{E}^{\mathbb{Q}}\left(\frac{B(t)}{B(T)}S(T)\mathbb{I}_{S(T)>K}\middle|\mathscr{F}_t\right) && \text{(B.16)} \\ &= \mathbb{E}^{\mathbb{Q}^S}\left(\frac{B(t)}{B(T)}S(T)\mathbb{I}_{S(T)>K}\mathbb{Z}_t\middle|\mathscr{F}_t\right) && \text{(B.17)} \\ &= S(t)\mathbb{E}^{\mathbb{Q}^S}\left(\mathbb{I}_{S(T)>K}|\mathscr{F}_t\right) && \text{(B.18)} \\ &= e^{X(t)}\mathbb{Q}^S(S(T)>K|\mathscr{F}_t) && \text{(B.19)} \\ &= e^{X(t)}P_1. && \text{(B.20)} \end{aligned}$$

This implies that the call price in (B.11) can be written in terms of both measures as

$$C(S(t),K,T-t,\Theta) = S(t)\mathbb{Q}^S(S(T)>K|\mathscr{F}_t) - Ke^{-r(T-t)}\mathbb{Q}(S(T)>K|\mathscr{F}_t). \quad \text{(B.21)}$$

Now, we shall compute the derivatives of $C \equiv C(S(t),K,\tau,\Omega)$ (which will be useful later):

$$\frac{\partial C}{\partial t} = e^{X(T)}\frac{\partial P_1}{\partial t} - Ke^{-r\tau}\left(rP_2 + \frac{\partial P_2}{\partial t}\right), \quad \text{(B.22)}$$

$$\frac{\partial C}{\partial X(t)} = e^{X(t)}\left(P_1 + \frac{\partial P_1}{\partial X(t)}\right) - Ke^{-r\tau}\frac{\partial P_2}{\partial X(t)}, \quad \text{(B.23)}$$

$$\frac{\partial^2 C}{\partial X^2(t)} = e^{X(t)}\left(P_1 + 2\frac{\partial P_1}{\partial X(t)} + \frac{\partial^2 P_1}{\partial^2 X(t)}\right) - Ke^{-r\tau}\frac{\partial^2 P_2}{\partial X^2(t)}, \quad \text{(B.24)}$$

$$\frac{\partial C}{\partial V(t)} = e^{X(t)}\frac{\partial P_1}{\partial V(t)} - Ke^{-r\tau}\frac{\partial P_2}{\partial V(t)}, \quad \text{(B.25)}$$

$$\frac{\partial^2 C}{\partial V^2(t)} = e^{X(t)}\frac{\partial^2 P_1}{\partial V^2(t)} - Ke^{-r\tau}\frac{\partial^2 P_2}{\partial V^2(t)} \quad \text{(B.26)}$$

and

$$\frac{\partial^2 C}{\partial X(t)\partial V(t)} = e^{X(t)}\left(\frac{\partial P_1}{\partial V(t)} + \frac{\partial^2 P_1}{\partial X(t)\partial V(t)}\right) - Ke^{-r\tau}\frac{\partial^2 P_2}{\partial X(t)\partial V(t)}. \quad \text{(B.27)}$$

However, we actually need the derivative with respect to $\tau = T-t$ rather than the derivative with respect to t. Using the chain rule, we obtain $\frac{\partial C}{\partial \tau} = \frac{\partial C}{\partial t}\frac{\partial t}{\partial \tau} = -\frac{\partial C}{\partial t}$.

Using the results demonstrated in Section 2.8, one could write the hedging portfolio PDE for the option C under Heston's framework as

$$-\frac{\partial C}{\partial \tau}+\frac{1}{2}V(t)\frac{\partial^2 C}{\partial X^2(t)}+\left(r-\frac{1}{2}V(t)\right)\frac{\partial C}{\partial X(t)}+\rho\sigma V(t)\frac{\partial^2 C}{\partial V(t)\partial X(t)} + \frac{1}{2}\sigma^2 V(t)\frac{\partial^2 C}{\partial V^2(t)}-rC+\kappa(\theta-V(t))\frac{\partial C}{\partial V(t)}=0. \quad \text{(B.28)}$$

Hence, using the derivatives computed earlier and some arduous algebra, one could get

$$e^{X(t)}\Gamma_1 - e^{-r\tau}K\Gamma_2 = 0 \quad \text{(B.29)}$$

where

$$\Gamma_j = -\frac{\partial P_j}{\partial \tau}+\rho\sigma V(t)\frac{\partial^2 P_j}{\partial X(t)\partial V(t)}+\frac{1}{2}V(t)\frac{\partial^2 P_j}{\partial X^2(t)}+\frac{1}{2}V(t)\sigma^2\frac{\partial^2 P_j}{\partial V^2(t)} + (r-u_j V(t))\frac{\partial P_j}{\partial X(t)}+(a-b_j V(t))\frac{\partial P_j}{\partial V(t)} \quad \text{(B.30)}$$

for $j=1,2$ and where $u_1=\frac{1}{2}$, $u_2=-\frac{1}{2}$, $a=\kappa\theta$, $b_1=\kappa-\rho\sigma$ and $b_2=\kappa$. Since (B.29) is true, we have no choice but to get

$$-\frac{\partial P_j}{\partial \tau}+\rho\sigma V(t)\frac{\partial^2 P_j}{\partial X(t)\partial V(t)}+\frac{1}{2}V(t)\frac{\partial^2 P_j}{\partial X^2(t)}+\frac{1}{2}V(t)\sigma^2\frac{\partial^2 P_j}{\partial V^2(t)} + (r-u_j V(t))\frac{\partial P_j}{\partial X(t)}+(a-b_j V(t))\frac{\partial P_j}{\partial V(t)}=0 \quad \text{(B.31)}$$

for $j=1,2$.

Then, when the characteristic functions $f_j(\phi,X(t),V(t),\tau)$ corresponding to the in-the-money probabilities P_1 and P_2 are known, each probability can be recovered from its characteristic function using Lévy Inversion Theorem:

$$P_j=\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathscr{R}\left(\frac{e^{-i\phi\log(K)}f_j(\phi,X(t),V(t),\tau)}{i\phi}\right)d\phi. \quad \text{(B.32)}$$

At maturity, the probabilities are subject to the terminal condition

$$P_j(X(T),V(T),0)=\mathbb{I}_{X(T)>K} \quad \text{(B.33)}$$

which means that if at expiry the option is in-the-money, then the probability is unity.

At this point, in order to get a solution, Heston made an assumption about the characteristic functions: he assumes that they are of the log-linear form

$$f_j(\phi, X(t), V(t), \tau) = \exp(C_j(\tau,\phi) + D_j(\tau,\phi)V(t) + \mathrm{i}\phi X(t)) \tag{B.34}$$

where C_j and D_j are coefficients. We also know, as a direct consequence of the Feymann-Kac theorem, that the characteristic functions f_j will follow the PDEs in (B.31). Virtually, this means that if a function $f(X(t),V(t),t)$ of the SDEs of $X(t)$ and $V(t)$ satisfies the PDE $\frac{\partial f}{\partial t} - rf + \mathscr{A}f = 0$ with terminal condition $f(X(T),V(T),T)$ in the Heston bivariate diffusion and $\mathscr{A}$ is the Heston generator of Section 2.8 such that

$$\mathscr{A} = rS\frac{\partial}{\partial S} + \frac{1}{2}VS^2\frac{\partial}{\partial S^2} + \kappa(\theta - V)\frac{\partial}{\partial V} + \frac{1}{2}\sigma^2 V\frac{\partial^2}{\partial V^2} + \rho\sigma VS\frac{\partial}{\partial V \partial S}, \tag{B.35}$$

then the solution to f is

$$f(X(t),V(t),t) = \mathbb{E}(f(\chi_T, T)|X(t) = X, V(t) = V). \tag{B.36}$$

Using $f(X(T),V(T),T),T) = e^{\mathrm{i}\phi X(T)}$ produces the solution

$$f(X(t),V(t),t) = \mathbb{E}\left(e^{\mathrm{i}\phi X(T)}|X(t) = X, V(t) = V\right) \tag{B.37}$$

which is the characteristic function for $X(T)$. Thus, the PDE for the latter function is, using the results in (B.31),

$$-\frac{\partial f_j}{\partial \tau} + \rho\sigma V(t)\frac{\partial^2 f_j}{\partial X(t)\partial V(t)} + \frac{1}{2}V(t)\frac{\partial^2 f_j}{\partial X^2(t)} + \frac{1}{2}V(t)\sigma^2\frac{\partial^2 f_j}{\partial V^2(t)} + (r + u_j V(t))\frac{\partial f_j}{\partial X(t)} + (a - b_j V(t))\frac{\partial f_j}{\partial V(t)} = 0. \tag{B.38}$$

To evaluate the PDEs of the caracteristic functions, we need the following derivatives

$$\frac{\partial f_j}{\partial \tau} = -\left(\frac{\partial C_j}{\partial \tau} + \frac{\partial D_j}{\partial \tau}\right) f_j, \tag{B.39}$$

$$\frac{\partial f_j}{\partial X(t)} = \mathrm{i}\phi f_j, \tag{B.40}$$

$$\frac{\partial^2 f_j}{\partial X^2(t)} = -\phi^2 f_j, \tag{B.41}$$

$$\frac{\partial f_j}{\partial V(t)} = D_j f_j, \tag{B.42}$$

$$\frac{\partial^2 f_j}{\partial V^2(t)} = D_j^2 f_j \tag{B.43}$$

and

$$\frac{\partial^2 f_j}{\partial X(t)\partial V(t)} = \mathrm{i}\phi D_j f_j. \tag{B.44}$$

By substituting these derivatives in (B.38), one could get

$$V(t)\left(-\frac{\partial C_j}{\partial \tau} + \rho\sigma\mathrm{i}\phi D_j - \frac{1}{2}\phi^2 + \frac{1}{2}\sigma^2 D_j^2 + u_j\mathrm{i}\phi - b_j D_j\right) - \frac{\partial C_j}{\partial \tau} + r\mathrm{i}\phi + aD_j = 0. \tag{B.45}$$

This produces two differential equations:

$$\frac{\partial D_j}{\partial \tau} = \rho\sigma\mathrm{i}\phi D_j - \frac{1}{2}\phi^2 + \frac{1}{2}\sigma^2 D_j^2 + u_j\mathrm{i}\phi - b_j D_j \tag{B.46}$$

and

$$\frac{\partial C_j}{\partial \tau} = r\mathrm{i}\phi + aD_j \tag{B.47}$$

with $C_j(0,\phi) = 0$ and $D_j(0,\phi) = 0$ as initial conditions.

Equation (B.46) is known as the Riccati equation (we refer the reader to Rouah (2011) for more detail on this matter) in D_j while the second ODE, (B.47), can be solved using straightforward integration once D_j is obtained.

Hence, using the solution of the Riccati equation, one could get

$$D_j = \frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{\sigma^2}\left(\frac{1 - e^{h_j\tau}}{1 - g_j e^{h_j\tau}}\right) \tag{B.48}$$

with

$$g_j = \frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{b_j - \rho\sigma\mathrm{i}\phi - h_j} \tag{B.49}$$

and

$$h_j = \sqrt{(\rho\sigma\mathrm{i}\phi - b_j)^2 - \sigma^2(2u_j\mathrm{i}\phi - \phi^2)}. \tag{B.50}$$

To obtain the solution of C_j, one could integrate the second equation, B.47. Hence,

$$C_j = \int_0^\tau r\mathrm{i}\phi dy + a\left(\frac{b_j - \rho\sigma\mathrm{i}\phi + h_j}{\sigma^2}\right)\int_0^\tau \left(\frac{1 - e^{h_j y}}{1 - g_j e^{h_j y}}\right) dy. \tag{B.51}$$

The first integral is $ri\phi\tau$ and the second one can be found using the substitution $x = e^{h_j y}$. Thus, we obtain

$$C_j = ri\phi\tau + \frac{a}{h_j}\left(\frac{b_j - \rho\sigma i\phi + h_j}{\sigma^2}\right)\int_0^{e^{h_j\tau}}\left(\frac{1-x}{1-g_j x}\right)\frac{1}{x}dx. \tag{B.52}$$

Using partial fractions to evaluate the integral, one could finally get

$$C_j = ri\phi r + \frac{a}{\sigma^2}\left((b_j - \rho\sigma i\phi)\tau - 2\log\left(\frac{1-g_j e^{h_j\tau}}{1-g_j}\right)\right) \tag{B.53}$$

which is the final solution for C_j.

This completes the original derivation of the European call price in the Heston framework. □

Appendix C

DISCUSSION ON THE NUMERICAL INTEGRATION TECHNIQUES OF PROPOSITION 2.4.1

Since the formulas given in Heston (1993) are semi-closed, we need a numerical integration method in order to get real numerical prices. One could use two criteria in order to select the right method: accuracy of the solution and the time it takes to get the solution. Essentially, in our discussion, we will consider quadrature rules since they allow us to quickly approximate the integrals of interest. Another approach would have been to use fast Fourier transforms; however, this technique has a bad reputation with probability estimation in the Heston framework.

An important number of researchers were interested in the numerical integration issues for the probabilities yielded by the Heston model. Janek *et al.* (2011) did a quick retrospective of what has been implemented so far. Hakala and Wystup (2002) propose a Gauss-Laguerre (we refer the reader to Hildebrand (1987) for more detail on numerical methods) quadrature rule using 100 as the upper bound of the integral (instead of infinity). Kahl and Jäckel (2006) suggest the use of the Gauss-Lobatto quadrature rule; they can do so by changing the integration interval from $[0,\infty)$ to $[0,1]$.

Finally, Janek *et al.* (2011) propose the use of the Gauss-Kronrod quadrature rule and demonstrate empirically the advantages of this method.

The Gauss-Kronrod is a variant of the Gaussian quadrature in which the evaluation points are chosen so that a more accurate estimate can be calculated using the same grid used for less accurate estimations. This kind of quadrature is very effective to obtain highly accurate results. In addition, this method yields good results when the integrand is oscillatory.

In light of this, we will use the Gauss-Kronrod quadrature rule for our numerical integrations.