



Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

**An advanced Monte Carlo method for the multi-asset
Heston model**

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by Wander Sybe Wadman

Delft, the Netherlands
November 3, 2010

Copyright © 2010 by Wander Sybe Wadman. All rights reserved.



MSc THESIS APPLIED MATHEMATICS

“An advanced Monte Carlo method for the multi-asset Heston model”

Wander Sybe Wadman

Delft University of Technology

Daily supervisor

Prof. Dr. Ir. C.W. Oosterlee

Ir. L.A. Grzelak

Other thesis committee members

Dr. J.A.M. van der Weide

Dr. H.M. Schuttelaars

Responsible professor

Prof. Dr. Ir. C.W. Oosterlee

Drs. W.F. (Erik) van Raaij

November 3, 2010

Delft, the Netherlands

Preface

As the final part of the two-year MSc-program Applied Mathematics at the Faculty of Electrical Engineering, Mathematics and Computer Science, students at Delft University of Technology can choose to perform their master's thesis in collaboration with a company which has a suitable mathematical topic of interest. The goal of this project is to design a new numerical method to price financial instruments for the Derivatives, Research & Validation department of Rabobank International. The time frame of the project is nine months.

I would like to thank my supervisor professor Cornelis Oosterlee for our fruitful discussions and my grateful use of his experience. At least as much thanks should go to Lech Grzelak, since his tireless support and inspiration for me to find new ideas were very helpful during the project. Furthermore, I would like to thank Steven Korteland for his attentive supervision during a part of the thesis. Last but not least, great love goes out to my girlfriend, who kept supporting me during my frequent absence, and to my parents, whose help gave me a lot of motivation.

Contents

1	Introduction	7
2	The Heston model	8
2.1	Analysis of the variance dynamics	9
2.2	Analysis of the stock dynamics	10
3	Path simulation of the variance process	12
3.1	Taylor-based schemes	13
3.1.1	Euler discretization: the Full Truncation scheme (FT)	13
3.1.2	Transformed Volatility scheme (TV)	14
3.2	(Almost) exact schemes	14
3.2.1	Exact simulation (ES)	15
3.2.2	Quadratic Exponential scheme (QE)	16
3.2.3	Noncentral Chi-square Inversion scheme (NCI)	17
3.2.4	Gamma Expansion scheme (GE)	18
3.3	Comparison of different schemes	19
4	The MQE Monte Carlo method	21
4.1	Assumptions	21
4.2	Construction of the MQE scheme	22
4.2.1	Set up of the MQE scheme	23
4.2.2	Simulation of the next time step	24
5	Performance tests for the MQE scheme	28
5.1	Imposed correlation test	28
5.1.1	Results of imposed correlation test	29
5.2	Option validation: the <i>Feller Satisfied</i> Test	30
5.3	Option validation: the <i>Feller Violated</i> Test I	31
5.4	Option validation: the <i>Feller Violated</i> Test II	32
5.4.1	Problem set up	33
5.4.2	Strategy outline in case of Black-Scholes dynamics	33
5.4.3	Change of numéraire under Heston	34
5.4.4	Dynamics of $S_q(t)$ under $\mathbb{S}$	35
5.4.5	Conditions for model affinity	37
5.4.6	Derivation of the characteristic function	39
5.4.7	Results	41
5.5	Conclusion	44
5.6	Further investigation	45
6	Calibration of the MQE scheme	47
6.1	Single-asset Heston calibration	48
6.2	Stock-stock correlation calibration	50
6.2.1	Literature on multi-asset Heston calibration	50
6.3	Market implied calibration method	51

6.3.1	Approximate second moment of S_s	52
6.3.2	Approximate distribution of $S_s(T)$	53
6.4	Investigation of the calibration method	54
6.5	Discussion of the calibration method	57
6.6	Conclusion and outlook	59
7	Conclusions	61
A	Cholesky decomposition	65
B	Derivation of the second moment of $S_s(T)$	67

Chapter 1

Introduction

The valuation of financial derivatives is becoming more and more a dedicated subject of mathematics and computer science. Nowadays, the prices of contracts whose future values are uncertain (foreign exchange, interest rate, equity, commodity) add up to billions of US dollars each year (see [BIS, 2009]). Because of this importance, several mathematical approaches have been developed to model the underlying of derivatives.

Some derivatives are based on more than one underlying. Examples are stock market index options or derivatives used to hedge against the fluctuation of more than one currency. As can be seen from historical market data, these underlyings are correlated in general. Therefore, pricing these derivatives would involve a multi-dimensional extension of the single asset models. This thesis considers the multi-asset Heston model.

The celebrated Heston model (see [Heston, 1993]) is an extension of the well-known Black-Scholes model (see [Black & Scholes, 1973]). An important drawback of the latter model is the strong assumption of a constant volatility. Market data analysis (as is done in [Cont, 2001], for example) indicates the existence of the ‘volatility skew’ or ‘smile’. The model of Heston allows such skew by defining the volatility as a stochastic process itself.

In this thesis, an extension of the Heston model to the multi-dimensional case will be investigated. Most attention will be given to design a multi-asset Monte Carlo method, which can efficiently simulate multivariate random variables with almost no bias. The prevention of negative variance in the discretization method will be a challenge, as well as the generation of correlated multivariate random variables.

In Chapter 2, the Heston model will be discussed. In Chapter 3, the problem of potential negative variance is treated. We investigate several numerical methods which are designed to simulate the variance path, and we test them. We will choose one method as our candidate scheme for the multi-dimensional Monte Carlo method. In Chapter 4, we construct this method, which we name the Multi-dimensional Quadratic Exponential method (MQE). We test its performance in Chapter 5. In Chapter 6, we derive a calibration method for the MQE method.

Chapter 2

The Heston model

Consider the probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, with $\mathbb{Q}$ the risk-neutral measure. The Heston model describes the $\mathbb{Q}$ -dynamics of a stock price, $S(t)$, with stochastic variance, $V(t)$, by the two-dimensional stochastic differential equation (SDE)

$$dV(t) = \kappa(\theta - V(t))dt + \varepsilon\sqrt{V(t)}dW_V(t), \quad (2.1)$$

$$dS(t) = r(t)S(t)dt + \sqrt{V(t)}S(t)dW_X(t). \quad (2.2)$$

Here the mean-reverting term, κ , the long-term variance, θ , and the volatility of volatility, ε , are strictly positive constants. The mean rate of return, $r(t)$, is actually assumed constant in the Heston model, but we assume that it is a deterministic function of time, since this natural generalization will not make our further derivations too complex. Further, $W_X(t)$ and $W_V(t)$ are Brownian motions under $\mathbb{Q}$, and we have $dW_X(t)dW_V(t) = \rho dt$, with dt an infinitesimal amount of time.

We will now define the concept of arbitrage and then use an important theorem in finance:

Definition. Arbitrage

An arbitrage is a portfolio value process $X(t)$ satisfying $X(0) = 0$ and also satisfying for some time $T > 0$

$$\mathbb{Q}\{X(T) \geq 0\} = 1, \quad \mathbb{Q}\{X(T) > 0\} > 0.$$

An arbitrage is a way of trading so that one starts with zero capital and some later time T is sure not to have lost money and furthermore has a positive probability of having earned money. We state the following well-known theorem without proof:

Theorem 1. First fundamental theorem of asset pricing.

If a market model has a risk-neutral probability, then it does not admit arbitrage.

Furthermore, since $\mathbb{Q}$ is the risk-neutral probability, $S(t)/B(t)$ is a martingale under $\mathbb{Q}$. Therefore, $r(t)$ must be equal to the risk-free rate of return, and we will refer to this by $r(t)$ from now on. For the multi-dimensional case, this will practically mean that the risk-free rate of every stock return is the same.

Since instantaneous correlation will be an important topic in this project, let us introduce this quantity.

Definition. Instantaneous correlation

Suppose $X_1(t)$ and $X_2(t)$ are stochastic processes with dynamics

$$dX_1(t) = \mu_1(t)dt + \sigma_1(t)dW_1(t),$$

$$dX_2(t) = \mu_2(t)dt + \sigma_2(t)dW_2(t),$$

where $W_1(t), W_2(t)$ are Brownian motions satisfying $dW_1(t)dW_2(t) = \rho(t)dt$ and $\rho(t), \mu_1(t), \mu_2(t), \sigma_1(t)$ and $\sigma_2(t)$ are functions of time. We call $\rho(t)$ the instantaneous correlation between $dX_1(t)$ and $dX_2(t)$.

Integration of the SDE of $X_i(t)$ for $i = 1, 2$ over a time interval $\Delta > 0$, yields

$$X_i(t + \Delta) - X_i(t) = \int_t^{t+\Delta} \mu_i(s) ds + \int_t^{t+\Delta} \sigma_i(s) dW_i(s).$$

We recognize the second integral on the right-hand side as an Itô integral, which is normally distributed given $\mathcal{F}(t)$, since $\sigma_i(s)$ is a deterministic function. Therefore:

$$X_i(t + \Delta) - X_i(t) \sim \mathcal{N} \left(\int_t^{t+\Delta} \mu_i(s) ds, \int_t^{t+\Delta} \sigma_i^2(s) dt \right) \text{ given } \mathcal{F}(t).$$

In this case of deterministic $\rho(t), \mu_1(t), \mu_2(t), \sigma_1(t)$ and $\sigma_2(t)$, the instantaneous correlation is equal to the correlation between the infinitesimal increases of the two stochastic processes. We can make this intuitive by assuming that these five parameters are constant: $\rho(t) = \rho, \mu_1(t) = \mu_1$, and so on. Then, for $t > 0$ and infinitesimal increase dt , it is easy to derive that

$$\begin{aligned} \mathbb{E}[X_i(t + dt) - X_i(t) | \mathcal{F}(t)] &= \mu_i dt, & \text{for } i = 1, 2, \\ \mathbb{V}\text{ar}(X_i(t + dt) - X_i(t) | \mathcal{F}(t)) &= \sigma_i^2 dt, & \text{for } i = 1, 2, \\ \text{Cov}(X_1(t + dt) - X_1(t), X_2(t + dt) - X_2(t) | \mathcal{F}(t)) &= \rho \sigma_1 \sigma_2 dt. \end{aligned}$$

Therefore, conditioned on $\mathcal{F}(t)$, the correlation between the increments $X_1(t + dt) - X_1(t)$ and $X_2(t + dt) - X_2(t)$ is

$$\frac{\text{Cov}(X_1(t + dt) - X_1(t), X_2(t + dt) - X_2(t) | \mathcal{F}(t))}{\sqrt{\mathbb{V}\text{ar}(X_1(t + dt) - X_1(t) | \mathcal{F}(t)) \mathbb{V}\text{ar}(X_2(t + dt) - X_2(t) | \mathcal{F}(t))}} = \rho.$$

If we drop the assumption of constant $\rho(t), \mu_1(t), \mu_2(t), \sigma_1(t)$ and $\sigma_2(t)$, a similar result can be derived. We present it here without proof (see [Shreve, 2004]):

$$\lim_{dt \downarrow 0} \frac{\text{Cov}(X_1(t + dt) - X_1(t), X_2(t + dt) - X_2(t) | \mathcal{F}(t))}{\sqrt{\mathbb{V}\text{ar}(X_1(t + dt) - X_1(t) | \mathcal{F}(t)) \mathbb{V}\text{ar}(X_2(t + dt) - X_2(t) | \mathcal{F}(t))}} = \rho(t) \text{ given } \mathcal{F}(t).$$

It is important to notice that this result will not hold for stochastic $\sigma_1(t)$ and $\sigma_2(t)$, as it is the case in the Heston model. In that case, we call $\rho(t)$ the instantaneous correlation between the Brownian motions of the Heston model, while the log-stock return correlation denotes the correlation between $dX_1(t)$ and $dX_2(t)$.

2.1 Analysis of the variance dynamics

The variance process in (2.1) is a mean-reverting square-root process. An investigation of its dynamics in [Cox et al., 1985] gives us the conditional distribution of the variance. We state some results without proof:

Result 1. Let $\Delta > 0$ and $F_{\chi^2}(x, \delta, \lambda)$ be the cumulative distribution function for the noncentral chi-square distribution with δ degrees of freedom and noncentrality parameter λ . Then

$$\mathbb{Q}(V(t + \Delta) < x | V(t)) = F_{\chi^2} \left(\frac{n(t, \Delta)x}{e^{-\kappa\Delta}}, d, n(t, \Delta)V(t) \right),$$

with $d = 4\kappa\theta/\varepsilon^2$ and $n(t, \Delta) = \frac{4\kappa e^{-\kappa\Delta}}{\varepsilon^2(1 - e^{-\kappa\Delta})}$.

Result 2. Since a noncentral chi-square distributed random variable, $X \sim F_{\chi^2}(x, k, \lambda)$, has mean $k + \lambda$ and variance $2(k + 2\lambda)$, we can compute the first two conditional moments of the variance process:

$$\begin{aligned} \mathbb{E}[V(t + \Delta) | V(t)] &= \theta + (V(t) - \theta)e^{-\kappa\Delta}, \\ \mathbb{V}\text{ar}[V(t + \Delta) | V(t)] &= \frac{V(t)\varepsilon^2}{\kappa} e^{-\kappa\Delta} (1 - e^{-\kappa\Delta}) + \frac{\theta\varepsilon^2}{2\kappa} (1 - e^{-\kappa\Delta})^2. \end{aligned}$$

2.2 Analysis of the stock dynamics

Let us define the log-stock, $X(t) := \ln S(t)$, and apply Itô's Lemma on the function $f(t, S(t)) = \ln S(t)$ to find the dynamics of $X(t)$:

$$dX(t) = \left(r(t) - \frac{1}{2}V(t)\right)dt + \sqrt{V(t)}dW_X(t). \quad (2.3)$$

By integrating over the time step, we obtain an expression which is intuitive to discretize:

$$X(t + \Delta) = X(t) + \int_t^{t+\Delta} \left(r(s) - \frac{1}{2}V(s)\right)ds + \int_t^{t+\Delta} \sqrt{V(s)}dW_X(s). \quad (2.4)$$

We know that we can sample a Brownian motion $dW(t)$ by a normal random variable with zero mean and variance dt . Furthermore, the correlation ρ between the two Brownian motions must be imposed. To do so, we can use a Cholesky decomposition.

Result 3. Cholesky decomposition

Suppose that an $n \times n$ matrix Σ is real-valued, symmetric and positive definite. Then we can write Σ as $LL^\top$, with L an $n \times n$ lower triangular matrix. This decomposition is called the Cholesky decomposition of Σ .

The proof of this theorem and an algorithm to find L can be found in Appendix A. We now claim that if $\mathbf{Z}$ is a vector of n independent standard normal random variables, then $L\mathbf{Z}$ is a vector of standard normal random variables with correlation matrix Σ . To show this, we first use the fact that the family of normal distributions is closed under linear transformations. Therefore, all elements of $L\mathbf{Z}$ are normally distributed. The expectation vector, μ_{LZ} , and covariance matrix, Σ_{LZ} , of $L\mathbf{Z}$ can easily be deduced, using the linear property of expectations:

$$\begin{aligned} \mu_{LZ} &= \mathbb{E}[L\mathbf{Z}] = L\mathbb{E}[\mathbf{Z}] = \mathbf{0} \\ \Sigma_{LZ} &= \mathbb{E}[(L\mathbf{Z} - \mathbb{E}[L\mathbf{Z}]) (L\mathbf{Z} - \mathbb{E}[L\mathbf{Z}])^\top] = \mathbb{E}[L\mathbf{Z} (L\mathbf{Z})^\top] = \mathbb{E}[L\mathbf{Z}\mathbf{Z}^\top L^\top] \\ &= L\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]L^\top = LIL^\top = \Sigma. \end{aligned}$$

In the second last equality, $\mathbb{E}[\mathbf{Z}\mathbf{Z}^\top]$ is equal to the identity matrix I since the components of $\mathbf{Z}$ are independent and standard normal random variables. Therefore, the multi-variate normal vector $\sqrt{\Delta}L\mathbf{Z}$ has correlation matrix Σ , but all elements still have mean zero and variance Δ .

If we define the 2-dimensional vector

$$d\mathbf{W}_{corr}(t) = \begin{pmatrix} dW_V(t) \\ dW_X(t) \end{pmatrix},$$

then its correlation matrix and its corresponding lower triangular matrix are

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}.$$

This implies that if

$$d\mathbf{W}_{uncorr}(t) = \begin{pmatrix} dW_V(t) \\ dW(t) \end{pmatrix},$$

with $dW(t)$ a Brownian motion independent of $dW_V(t)$, then $Ld\mathbf{W}_{uncorr}(t)$ has correlation matrix Σ . This practically means that in our scheme we have to sample two independent Brownian motions $dW_V(t)$ and $dW(t)$, and set

$$dW_X(t) = \rho dW_V(t) + \sqrt{1-\rho^2}dW(t). \quad (2.5)$$

Suppose we do not sample the numerical estimate of $dW_V(t)$ itself to simulate the variance process, but another random variable instead (for example the noncentral chi-square random variables as defined in Section 2.1). Then we cannot

use the above procedure to correlate the two processes. Although there are methods to correlate non-normal random variables (for example, by sampling from the normal copula), these are not exact, unfortunately. In [Broadie & Kaya, 2006], the following solution is proposed: first, substitute (2.5) in (2.4):

$$X(t + \Delta) = X(t) + \int_t^{t+\Delta} \left(r(s) - \frac{1}{2} V(s) \right) ds + \rho \int_t^{t+\Delta} \sqrt{V(s)} dW_V(s) + \sqrt{1 - \rho^2} \int_t^{t+\Delta} \sqrt{V(s)} dW(s). \quad (2.6)$$

Then we can replace $dW_V(t)$ by terms that we do sample, namely $V(t)$ and $V(t + \Delta)$. Therefore, we integrate the variance process SDE, formula (2.1):

$$V(t + \Delta) - V(t) = \int_t^{t+\Delta} \kappa(\theta - V(s)) ds + \varepsilon \int_t^{t+\Delta} \sqrt{V(s)} dW_V(s).$$

Rearranging yields

$$\int_t^{t+\Delta} \sqrt{V(s)} dW_V(s) = \varepsilon^{-1} \left(V(t + \Delta) - V(t) - \kappa\theta\Delta + \kappa \int_t^{t+\Delta} V(s) ds \right). \quad (2.7)$$

Finally, the substitution of (2.7) in (2.6) yields

$$\begin{aligned} X(t + \Delta) = & \underbrace{X(t)}_{v_1} + \underbrace{\int_t^{t+\Delta} r(s) ds}_{v_2} - \underbrace{\frac{1}{2} \int_t^{t+\Delta} V(s) ds}_{v_3} + \underbrace{\frac{\rho}{\varepsilon} (V(t + \Delta) - V(t))}_{v_4} \\ & + \underbrace{\frac{\kappa\rho}{\varepsilon} \int_t^{t+\Delta} (V(s) - \theta) ds}_{v_5} + \underbrace{\sqrt{1 - \rho^2} \int_t^{t+\Delta} \sqrt{V(s)} dW(s)}_{v_6}. \end{aligned} \quad (2.8)$$

The first term on the right-hand side is of course the current log-stock value. The second term represents the contribution of the interest rate during the time step. The third term is due to the quadratic variation of the geometric Brownian motion. The fourth term reflects the correlation between the two SDEs: in case of a positive ρ , an increase or decrease of the variance implies that this term respectively increases or decreases the stock value. (In case of a negative ρ , an increase or decrease of the variance implies that this term respectively decreases or increases the stock value.) The fifth term corresponds to this correlation in the same way, but now the *variation of the variance path from the long-term variance* is the driving factor, instead of the increase in variance. The last term represents the independent random part of the next log-stock value. Simplifying (2.8), we end up with the following exact representation of the log-stock process:

$$\begin{aligned} X(t + \Delta) = & X(t) + \int_t^{t+\Delta} r(s) ds + \frac{\rho}{\varepsilon} (V(t + \Delta) - V(t) - \kappa\theta\Delta) \\ & + \left(\frac{\kappa\rho}{\varepsilon} - \frac{1}{2} \right) \int_t^{t+\Delta} V(s) ds + \sqrt{1 - \rho^2} \int_t^{t+\Delta} \sqrt{V(s)} dW(s). \end{aligned} \quad (2.9)$$

This formula practically implies the following sampling scheme:

1. Sample the next step variance, $V(t + \Delta)$, and sample or estimate the variance path integral, $\int_t^{t+\Delta} V(s) ds$. In Chapter 3, we will describe several methods to do this.
2. Since $r(t)$ is deterministic, we can obtain $\int_t^{t+\Delta} r(s) ds$ independent of this sampling scheme.
3. Given $V(t)$, the last term of the right-hand side of (2.9) is an Itô integral, and we can thus sample this last term by drawing a normal random variable with mean zero and variance $(1 - \rho^2) \int_t^{t+\Delta} V(s) ds$.

Since the current log-stock value $X(t)$ is given, we have now obtained all terms on the right-hand side of (2.9), so we have obtained a sample of the next log-stock value $X(t + \Delta)$.

Chapter 3

Path simulation of the variance process

Let us take a closer look at the variance dynamics:

$$dV(t) = \kappa(\theta - V(t))dt + \varepsilon\sqrt{V(t)}dW_V(t), \quad (3.1)$$

We know from Result 1 in Section 2.1 that the next step variance, $V(t + \Delta)$, follows a noncentral chi-square distribution given the current step variance, $V(t)$. Since this distribution is supported on $(0, \infty)$, the variance process is non-negative for all t .

Secondly, observe that for larger values of ε , the variance process becomes more volatile. One may wonder whether the mean-reverting term, κ , can prevent the variance process to become zero. In [Feller, 1951], this phenomenon is formalized: precisely when $\varepsilon^2 < 2\kappa\theta$, then $P(V(t) = 0) = 0 \ \forall t > 0$. This condition is called the Feller condition, and it basically means that the variance process can not reach zero when the mean-reverting term is sufficiently strong.

The Feller condition is important for the choice of the discretization method. Suppose for example that we would discretize (3.1) by use of the well-known Forward Euler method:

$$\hat{V}(t + \Delta) = \hat{V}(t) + \kappa(\theta - \hat{V}(t))\Delta + \varepsilon\sqrt{\hat{V}(t)}\sqrt{\Delta}Z_V, \quad (3.2)$$

with Z_V a standard normal random variable. Then conditional on some non-zero current variance estimate $\hat{V}(t)$, the probability that the variance scheme is negative at the next step is

$$\begin{aligned} P(\hat{V}(t + \Delta) < 0 | \hat{V}(t)) &= P\left(Z_V < \frac{\kappa(\hat{V}(t) - \theta)\Delta - \hat{V}(t)}{\varepsilon\sqrt{\Delta\hat{V}(t)}} \middle| \hat{V}(t)\right) \\ &= \Phi\left(\frac{\kappa(\hat{V}(t) - \theta)\Delta - \hat{V}(t)}{\varepsilon\sqrt{\Delta\hat{V}(t)}}\right), \end{aligned} \quad (3.3)$$

with $\Phi(x)$ the cumulative distribution function of a standard normal random variable. The right-hand side of (3.3) will be strictly positive since $\hat{V}(t), \Delta$ and all Heston parameters are finite and non-zero. This means that with Euler Forward, there is a strictly positive probability that the discretization path becomes negative. This is undesirable, since it contradicts with the non-negative variance dynamics and it breaks down the scheme. In the subsequent step namely, the square-root of the negative variance is used, which is not a real number. Note that the probability of a breakdown in the computation of (3.3) increases when Δ increases, i.e. when the discretization is less accurate. Large values of ε , which correspond to volatile variance processes, imply a high probability as well. One can see this by writing (3.3) as

$$P(\hat{V}(t + \Delta) < 0 | \hat{V}(t)) = \Phi\left(\frac{\hat{V}(t)(\kappa\Delta - 1) - \kappa\theta\Delta}{\varepsilon\sqrt{\Delta\hat{V}(t)}}\right). \quad (3.4)$$

Since the numerator of the above fraction is negative for all practical parameter values, the probability increases in ε . In fact, the probability approaches $\Phi(0) = 0.5$ as ε approaches infinity. This positive probability of negative variance illustrates the importance of the Feller condition.

The question arises whether or not the Feller condition is violated in practice. The answer is that for typical examples of FX options, interest rate options as well as equity options, this is the case most of the time. Therefore, during the last decade, several numerical schemes have been proposed to overcome the problem of negative variance. In the first two sections of this chapter, we will introduce some of them, give an argumentation for the approximation used in the scheme and discuss possible drawbacks. In the last section of this chapter, we will compare their accuracy and speed in the one-dimensional case, as well as their applicability for the multi-dimensional case. Then we will decide which scheme will be the candidate scheme for the multi-dimensional Monte Carlo method that we will design.

3.1 Taylor-based schemes

We can distinguish two main categories of variance process discretization schemes: Taylor-based schemes and (almost) exact schemes. Taylor-based schemes use a discretization method with a constant estimate for the variance during each time step, taking the current variance for example. Since this constant estimate contains an error, one can increase accuracy by decreasing the time step.

We will discuss the Full Truncation scheme and the Transformed Volatility (TV) scheme. Full Truncation is one of the various Euler Forward schemes, but we will only discuss this scheme since in [Lord et al., 2008], it has been shown that it contains the least discretization bias.

3.1.1 Euler discretization: the Full Truncation scheme (FT)

To prevent $\hat{V}(t + \Delta)$ in (3.2) to become negative, one technique is to project any negative variance to zero. Since the variance process remains non-negative in this way, we prevent the computation of the square-root of a negative value. Instead of projection, reflection (changing the sign of a value) could be used for the same purpose. Furthermore, instead of *preventing* the variance process to become negative, one also could project or reflect negative variance values *before* they are used in further computations. Then the variance path itself can become negative, but before square-root computations (and maybe other computations as well) are performed with these negative values, projection or reflection is used.

The Euler FT scheme is based on this last method, using projection: the variance process is allowed to become negative, at which point the scheme becomes deterministic with a mean-reverting drift of $\kappa\theta$. As a result, for negative variance values, the scheme reduces to a mean reverting scheme. The algorithm is as follows:

$$\begin{aligned}\hat{V}(t + \Delta) &= \hat{V}(t) + \kappa(\theta - \hat{V}(t)^+) \Delta + \varepsilon \sqrt{\hat{V}(t)^+} \sqrt{\Delta} Z_V, \\ \hat{X}(t + \Delta) &= \hat{X}(t) + \left(r(t) - \frac{1}{2} \hat{V}(t)^+\right) \Delta + \sqrt{\hat{V}(t)^+} \sqrt{\Delta} Z_X,\end{aligned}$$

where $\hat{V}(t)$ and $\hat{X}(t)$ are the discrete estimates of $V(t)$ and $X(t)$ respectively, and $x^+ := \max(0, x)$.

The scheme is intuitive and very fast per time step. However, since the variance during $[t, t + \Delta]$ is estimated by the constant $\hat{V}(t)$, this estimation contains bias which increases in the time step size. Therefore, the accuracy of the method decreases in Δ , which is a well-known feature of the Euler Forward scheme. This means that we need a small time step to obtain sufficient accuracy. This bias increases whenever projection is applied more frequently, which is why this problem is significant when the Feller condition is violated or close to violation. In this case, the time step size must be so small that the scheme is often too slow.

3.1.2 Transformed Volatility scheme (TV)

In [Zhu, 2008], the author tries to prevent the negative variance in a different way: the idea of the TV scheme is to discretize the volatility dynamics instead of the variance dynamics. It thereby avoids the computation of the square-root of the process. To be precise, Zhu defines the volatility process $v(t)$ so that $V(t) = v^2(t)$ and applies Itô's lemma with $f(V(t), t) := \sqrt{V(t)}$ on formula (3.1), to obtain

$$\begin{aligned} dv(t) := d\sqrt{V(t)} &= \frac{1}{2}\kappa\left[\left(\theta - \frac{\varepsilon^2}{4\kappa}\right)v^{-1}(t) - v(t)\right]dt + \frac{1}{2}\varepsilon dW_V(t), \\ &= \kappa_v[\theta_v(t) - v(t)]dt + \varepsilon_v dW_V(t). \end{aligned} \quad (3.5)$$

Note that the volatility process can become negative, but this does not cause any unnatural behavior of the stock process:

$$dX(t) = \left(r(t) - \frac{1}{2}V(t)\right)dt + v(t)dW_X(t).$$

One can see that a change of sign of the volatility, $v(t)$, does not change the distribution of $X(t)$, by symmetry of $W_X(t)$.

Unfortunately, the application of Itô's lemma is not allowed here, since the derivatives

$$\frac{df}{dv}(v) = \frac{1}{2\sqrt{v}} \text{ and } \frac{d^2f}{dv^2}(v) = -\frac{1}{4v\sqrt{v}}$$

are obviously not continuous in $v = 0$. This means that the obtained model is not equivalent to the original Heston model. To illustrate this further, note that in [Heston, 1993], the mean-reverting square-root model (2.1) is derived from the Ornstein-Uhlenbeck process $dv(t) = -\frac{\kappa}{2}v(t)dt + \frac{\varepsilon}{2}dW_V(t)$, by Itô's lemma with $g(v(t), t) = v^2(t)$. However, process (3.5) is not this Ornstein-Uhlenbeck process again, since θ_v is stochastic.

Furthermore, the mean-reverting function of the first term disappears whenever $\theta < \frac{\varepsilon^2}{4\kappa}$. Then, the term will drive the volatility towards negative infinity whenever $v(t) > 0$, and towards positive infinity whenever $v(t) < 0$. Zhu proposes two techniques to mute these oscillations, one with a predictor-corrector method, and one by moment matching. The scheme algorithm is as follows:

$$\begin{aligned} \hat{X}(t + \Delta) &= \hat{X}(t) + \left(r(t) - \frac{1}{2}\hat{v}^2(t)\right)\Delta + \hat{v}(t)\sqrt{\Delta}Z_X, \\ \hat{v}(t + \Delta) &= \hat{v}(t) + \frac{1}{2}\kappa[\theta_v(t) - \hat{v}(t)]\Delta + \frac{1}{2}\varepsilon\sqrt{\Delta}Z_v, \end{aligned}$$

where $\hat{v}(t)$ is the discretization estimate of $v(t)$.

In [Zhu, 2008], it is proposed to use this model instead of the square-root process, since then potential negative volatilities do not cause any problems. However, if $\theta < \frac{\varepsilon^2}{4\kappa}$ (which is actually stronger than the violation of the Feller condition), the variance process is indeed not reverting to its mean anymore. The two proposed fixes do not solve this problem sufficiently for certain (in practice not uncommon) parameter sets. In all other cases, the TV scheme seems to be approximately as fast and accurate as the Euler FT scheme.

3.2 (Almost) exact schemes

A scheme that is based on sampling from the exact conditional distribution of $V(t + \Delta)$ is called an exact scheme. Since samples are exact, time steps do not have to be small to obtain sufficient accuracy. As we will show in the next subsection, it appears that the exact scheme for the variance path is computationally very expensive. Therefore, “almost exact” schemes have been proposed, which are, each in their own way, approximations of the exact scheme. For most schemes, a refinement of the time discretization will reduce the error of these approximations. However, an almost exact scheme will only be of use if it has the same or improved accuracy compared to the Euler FT scheme with respect to CPU time.

3.2.1 Exact simulation (ES)

In the well-known article [Broadie & Kaya, 2006], the authors introduced an alternative sampling scheme for the Heston model. The method of [Broadie & Kaya, 2006] is computationally expensive, which is why it will not be a serious candidate for the multi-dimensional Heston model. However, since it influenced several other methods developed afterwards, we will discuss it here briefly. The scheme algorithm is as follows:

1. Firstly, $V(t + \Delta)$ given $V(t)$ is sampled from a noncentral chi-square distribution, (see Result 1 of Chapter 2). The following representation for this distribution is used:

$$\chi'^2(d, \lambda) \stackrel{d}{=} \begin{cases} (Z + \sqrt{\lambda})^2 + \chi_{d-1}^2 & \text{for } d > 1, \\ \chi_{d+2Y_1}^2 & \text{for } d > 0. \end{cases}$$

with $Z \sim \mathcal{N}(0, 1)$, $\chi'^2(d, \lambda)$ the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter λ ; χ_{d+2N}^2 is the ordinary chi-square distribution with d degrees of freedom and Y_1 is Poisson distributed with mean $\frac{1}{2}\lambda$. In both cases, we have to sample from the ordinary chi-square distribution, which is a special case of the gamma distribution: $\chi_k^2 \sim \Gamma(k/2, 2)$. We can efficiently sample from $\Gamma(k/2, 2)$ by use of the following acceptance-rejection method:

- (a) We sample an exponentially distributed random variable, $Y_2 \sim \text{Exp}(k^{-1})$, by setting $Y_2 = -k \ln(1 - U_1)$, with $U_1 \sim U(0, 1)$. We sample a uniform random variable, U_2 , independent of Y_2 .
- (b) We define $f_\Gamma(x, k/2, 2)$ as the probability density function of a gamma random variable with shape $k/2$ and scale 2, and $g_{Y_2}(t)$ the probability density function of Y_2 . Then if

$$U_2 \leq \frac{f_\Gamma(Y_2, k/2, 2)}{g_{Y_2}(Y_2)},$$

we set $Y_3 = Y_2$, and $Y_3 \sim \Gamma(k/2, 2)$. Otherwise, we return to step (a).

2. Then, $\int_t^{t+\Delta} V(s) ds$ is sampled conditioned on starting point $V(t)$ and endpoint $V(t + \Delta)$. Its cumulative distribution function is recovered by a Fourier inversion method of the characteristic function $\phi(u)$ of $\int_t^{t+\Delta} V(s) ds$:

$$\begin{aligned} F(x) := P\left(\int_t^{t+\Delta} V(s) ds \leq x\right) &= \frac{2}{\pi} \int_0^\infty \frac{\sin(ux)}{u} \Re(\phi(u)) du \\ &\approx \frac{hx}{\pi} + \frac{2}{\pi} \sum_{j=1}^M \frac{\sin(hjx)}{j} \Re(\phi(hj)), \end{aligned} \quad (3.6)$$

where the Trapezoidal rule is used with step size h and M is chosen large. A numerical inversion of the CDF allows to sample $\int_t^{t+\Delta} V(s) ds$ by $F^{-1}(U)$, with U a uniform random number in $[0, 1]$.

3. Via (2.9), $X(t + \Delta)$ is sampled given $V(t)$, $V(t + \Delta)$ and $\int_t^{t+\Delta} V(s) ds$, using the knowledge that $\int_t^{t+\Delta} \sqrt{V(s)} dW(s)$ is normally distributed with mean zero and variance $\int_t^{t+\Delta} V(s) ds$.
4. We set $S(t + \Delta) = e^{X(t + \Delta)}$.

The simulation is exact except that the truncation of the sum and the numerical inversion of the CDF in (3.6) will introduce some bias. By choosing M large enough and the discretization step h small enough, we can achieve any desirable accuracy of the numerical inversion, and thus this bias can be as small as we desire. However, the characteristic

function contains two modified Bessel functions of the first kind:

$$\begin{aligned}\phi(x) &= \frac{\gamma(x)e^{-0.5\Delta(\gamma(x)-\kappa)}(1-e^{-\Delta\kappa})}{\kappa(1-e^{-\Delta\gamma(x)})} \\ &\times \exp \left[\frac{V(t)+V(t+\Delta)}{\varepsilon^2} \left(\frac{\kappa(1+e^{-\Delta\kappa})}{1-e^{-\kappa\Delta}} - \frac{\gamma(x)(1+e^{-\Delta\gamma(x)})}{1-e^{-\Delta\gamma(x)}} \right) \right] \\ &\times \frac{I_{0.5d-1} \left[\sqrt{V(t)V(t+\Delta)} \frac{4\gamma(x)e^{-0.5\Delta\gamma(x)}}{\varepsilon^2(1-e^{-\Delta\gamma(x)})} \right]}{I_{0.5d-1} \left[\sqrt{V(t)V(t+\Delta)} \frac{4\kappa e^{-0.5\Delta\kappa}}{\varepsilon^2(1-e^{-\Delta\kappa})} \right]},\end{aligned}$$

with $\gamma(x) = \sqrt{\kappa^2 - 2\varepsilon^2 ix}$, $d = \frac{4\kappa\theta}{\varepsilon^2}$ and $I_\nu(x)$ is the modified Bessel function of the first kind. The method becomes very time consuming because of the computation of $2M$ of these Bessel functions each step.

Furthermore, the acceptance-rejection method has the disadvantage that the number of samples depend on the specific Heston parameter set, which influences the quasi randomness of the random number generator. The methods to follow in subsequent sections will address some of these problems by adapting the method in [Broadie & Kaya, 2006] in some way.

3.2.2 Quadratic Exponential scheme (QE)

Recall from Result 1 in Section 2.1 that

$$\frac{n(t, \Delta)x}{e^{-\kappa\Delta}} V(t + \Delta) \sim \chi'^2(d, n(t, \Delta)V(t)).$$

It was observed in [Andersen, 2007] that a noncentral chi-square random variable $X \sim \chi'^2(d, \lambda)$ approaches a normal distribution as λ approaches infinity. Therefore, Andersen proposed to estimate the next step variance by a function of a normal random variable. The moments can be matched using Result 2. Observe that a normal random variable can attain negative values, whereas $V(t + \Delta)$ can not. Therefore, Andersen proposed to either truncate or square the normal random variable. The former scheme, the Truncated Gaussian scheme (TG), is slightly outperformed by the latter scheme, the Quadratic Exponential scheme (QE). We will only discuss the QE scheme for this reason. In this scheme, $V(t + \Delta)$, with a moderate or high noncentrality parameter, is estimated by:

$$\hat{V}(t + \Delta) = a(b + Z_V)^2, \quad (3.7)$$

with Z_V a standard normal random variable. For $\lambda \downarrow 0$ however, this approximation becomes inaccurate, but then $\chi'^2(d, \lambda)$ approaches a *central* chi-square distribution, $\chi^2(d)$. The corresponding probability density function reads

$$f(x) = \frac{e^{-y/2} y^{d/2-1}}{2^{d/2} \Gamma(d/2)}.$$

with $d = \frac{4\kappa\theta}{\varepsilon^2}$ degrees of freedom. Based on the arguments given at the end of Section 3.2.1, Andersen proposed to approximate this distribution too. One can see from the PDF that when the Feller condition is violated, then d is small, so then the density of $V(t + \Delta)$ will be relatively large around 0. This insight gives rise to an approximation of the PDF by a distribution with some point mass at the origin, supplemented with an exponential tail:

$$\Psi(x, p, \beta) = P(\hat{V}(t + \Delta) \leq x) = p + (1 - p)(1 - e^{-\beta x}).$$

Its inverse, $\Psi^{-1}(x, p, \beta)$ exists in explicit form. Therefore, can sample exactly from this distribution and set our new value for the variance

$$\hat{V}(t + \Delta) = \Psi^{-1}(U_V, p, \beta) = \begin{cases} 0 & \text{for } 0 \leq U_V \leq p, \\ \beta^{-1} \ln \left(\frac{1-p}{1-U_V} \right) & \text{for } p < U_V \leq 1, \end{cases} \quad (3.8)$$

with U_V a uniform random variable. Constants a and b from (3.7) and p and β from (3.8) are determined by matching the moments of this distribution with those of the true conditional distribution of $V(t + \Delta)$. The scheme algorithm is as follows:

1. To find a rule to switch between the two estimates for $V(t + \Delta)$, we first define

$$\Psi = \frac{\mathbb{V}\text{ar}[V(t + \Delta)|V(t) = \hat{V}(t)]}{\mathbb{E}[V(t + \Delta)|V(t) = \hat{V}(t)]^2}.$$

For the simulation of $\hat{V}(t + \Delta)$, we use (3.7) when $\Psi \leq \Psi_c$, with $\Psi_c \in [1, 2]$ the critical value. Whenever $\Psi > \Psi_c$, we use (3.8). Andersen claims that the precise choice of Ψ_c does not matter for the accuracy of the scheme, and chooses $\Psi_c = 1.5$.

2. Then $\int_t^{t+\Delta} V(s)ds$ is estimated using $\int_t^{t+\Delta} V(s)ds \approx \Delta[\gamma_1 V(t) + \gamma_2 V(t + \Delta)]$, with, for example, $\gamma_1 = \gamma_2 = \frac{1}{2}$, or by matching moments.
3. Finally, $\hat{X}(t + \Delta)$ is generated using (2.9), just as in the exact scheme as described in Section 3.2.1.

The QE method performs very fast, since it uses just few more computations per time step than the Euler FT scheme. Its accuracy is superior to that of the other schemes, especially when the Feller condition is violated. This is due to the fact that the method is based on the exact distribution of the next step variance, and not on its Taylor expansion.

3.2.3 Noncentral Chi-square Inversion scheme (NCI)

Let us recall from Section 3.2.1, that a noncentral chi-square distribution $\chi'^2(d, \lambda)$ can be represented by a central chi-square distribution $\chi_{d+2Y_1}^2$ with stochastic degrees of freedom: Y_1 is Poisson distributed with mean $\frac{1}{2}\lambda$. Applying this to the conditional sampling method of $V(t + \Delta)$, implies drawing a Poisson number and an inversion of the ordinary chi-square CDF. While the exact simulation in [Broadie & Kaya, 2006] uses an acceptance-rejection test (see Section 3.2.1) instead of the CDF inversion, in [van Haastrecht & Pelsner, 2008], it is claimed that this method is too slow. However, inverting the ordinary chi-square CDF each step and each path is numerically expensive. Therefore, the authors of [van Haastrecht & Pelsner, 2008] propose to use an interpolation on some cached values. The scheme algorithm is as follows:

1. Based on the Poisson mean $\frac{1}{2}\lambda$, a integer grid $\mathcal{M} := \{0, \dots, M_j, \dots, M_{\max}\}$ is chosen as well as a grid $\mathcal{U}_{M_j} := \{0, \dots, 1 - \delta\}$, with δ some small number. $\mathcal{M}$ and $\mathcal{U}_{M_j}$ contain the Poisson values and the discretized domain of the inverse, respectively, on which we will cache the inverse of the corresponding conditional chi-square CDF's:

$$H_{M_j}^{-1}(U_i) := G_{\chi_{d+2M_j}^2}^{-1}(U_i), \quad \forall M_j \in \mathcal{M} \quad \forall U_i \in \mathcal{U}_{M_j},$$

with $G_{\chi_{d+2M_j}^2}^{-1}(x)$ the inverse of the chi-square CDF with $d + 2M_j$ degrees of freedom. Note that we chose $1 - \delta$ instead of 1 as the final grid point of $\mathcal{U}_{M_j}$ to omit the infinite value $G_{\chi_{d+2M_j}^2}^{-1}(1)$ for all M_j .

2. We draw a Poisson distributed random variable, m_j^1 . Now we can sample $V(t + \Delta|V(t))$ by drawing a uniform random number U_V and by computing

$$V(t + \Delta) = \begin{cases} e^{-\kappa\Delta} J(U_V)/n(t, \Delta) & \text{for } m_j \leq M_{\max}, \\ e^{-\kappa\Delta} F_{d+2M_j}^{-1}/n(t, \Delta) & \text{for } m_j > M_{\max}, \end{cases} \quad (3.9)$$

with $n(t, \Delta)$ as in Proposition 1, $J(\cdot)$ an interpolation rule on $H_{M_j}^{-1}$, and $F_{d+2M_j}^{-1}$ the direct numerical inversion of the ordinary chi-square CDF. The authors of [van Haastrecht & Pelsner, 2008] propose to use a monotone cubic Hermite spline interpolation for $J(\cdot)$.

3. $\int_t^{t+\Delta} V(s) ds$ and $X(t + \Delta)$ are generated just as in the QE scheme.

¹In [Ahrens & Dieter, 1982] for example, an efficient method is described

The idea is to choose $\mathcal{M}$ in such a way that the caching does not cost too much computational effort, while the interpolation in (3.9) must be used often enough to speed up the original method of direct inversion. Furthermore, $\mathcal{U}_{M_j}$ must be chosen so that the interpolation is accurate enough. For example, both grids could be made non-equidistant to improve the accuracy.

In case the Feller condition is violated, the NCI scheme performs well. In [van Haastrecht & Pelsner, 2008], it is claimed that $\mathbb{E}[N] = \frac{1}{2}\lambda \ll 10$ in all practical model configurations, which would make the Poisson generator computationally efficient. However, in case of more “Black Scholes type” parameter settings (id est, when ε is small with respect to $\kappa\theta$), $\mathbb{E}[\lambda]$ can become much larger than 10. Since the Poisson random number generator employs an acceptance-rejection method, each natural number will be rejected until the first acceptance, which will slow down the scheme for high λ . In this case, we should switch from the NCI scheme to another method, for example QE, as proposed in [van Haastrecht & Pelsner, 2008]. Anyhow, the NCI scheme can not be a generic method for the Heston model by this argument.

3.2.4 Gamma Expansion scheme (GE)

While in [Andersen, 2007] and [van Haastrecht & Pelsner, 2008], the aim is to speed up the generation of $V(t + \Delta)$, the authors of [Glasserman & Kim, 2008] sample $\int_t^{t+\Delta} V(s)ds$ (given $V(t)$ and $V(t + \Delta)$) more efficiently than is done in the exact scheme. The article proposes an approximation of its distribution that should be accurate enough compared to the ones in the QE scheme or the NCI scheme. This approximation is based on the following proposition, proved in the article.

Proposition 1.

$$\left(\int_t^{t+\Delta} V(s)ds | V(t) = v_t, V(t+\Delta) = v_{t+\Delta} \right) \stackrel{d}{=} X_1 + X_2 + \sum_{j=1}^{\eta} Z_j, \quad (3.10)$$

in which $X_1, X_2, \eta, Z_1, Z_2, \dots$ are mutually independent random variables. The following representations hold:

$$X_1 \stackrel{d}{=} \sum_{n=1}^{\infty} \frac{1}{\gamma_n} \sum_{j=1}^{N_n} \text{Exp}_j(1), \quad X_2 \stackrel{d}{=} \sum_{n=1}^{\infty} \frac{1}{\gamma_n} \Gamma_n(d/2, 1), \quad Z_j \stackrel{d}{=} \sum_{n=1}^{\infty} \frac{1}{\gamma_n} \sum_{j=1}^{N_n} \Gamma_n(2, 1), \quad \forall j = 1, 2, \dots, \quad (3.11)$$

where

$$d = 4\kappa\theta/\varepsilon^2, \quad \lambda_n = \frac{16\pi^2 n^2}{\varepsilon^2 t(\kappa^2 t^2 + 4\pi^2 n^2)}, \quad \gamma_n = \frac{\kappa^2 t^2 + 4\pi^2 n^2}{2\varepsilon^2 t^2},$$

the Y_n are independent Poisson random variables with respective means $(v(t) + v(t + \Delta))\lambda_n$, the $\text{Exp}_j(1)$ are independent exponential random variables with mean 1, and the $\Gamma_n(\alpha, \beta)$ are independent gamma random variables with shape parameter α and scale parameter β . η is a Bessel random variable, which has probability mass function

$$P(\eta = n) = \frac{(z/2)^{2n+v}}{I_v(z)n!\Gamma(n+v+1)}, \quad n \geq 0,$$

with parameters $v = d/2 - 1$, $I_v(z)$ a modified Bessel function of the first kind and

$$z = \frac{2\kappa/\varepsilon^2}{\sinh(\kappa t/2)} \sqrt{v_t v_{t+\Delta}}.$$

For the precise scheme algorithm, see [Glasserman & Kim, 2008]. The idea of the sampling method of the integral in (3.10) is to truncate the infinite sums in (3.11) and match the moments of the truncation error estimate.

Just as the NCI scheme, this method performs well for the extreme cases that are treated in the article. That is, in case the Feller condition is violated. However, again for $\varepsilon \downarrow 0$, the computational speed decreases enormously, this time because the computation of the Bessel function values is consuming since its parameter z has become large. We conclude that just as the NCI scheme, the GE scheme cannot become a generic scheme for the Heston model.

3.3 Comparison of different schemes

We implemented the 5 schemes described in the above subsections in MatLab 2007b. To show that the schemes can perform well in at least one particular case, the European call option price estimates of these schemes are shown in Figure 3.1.

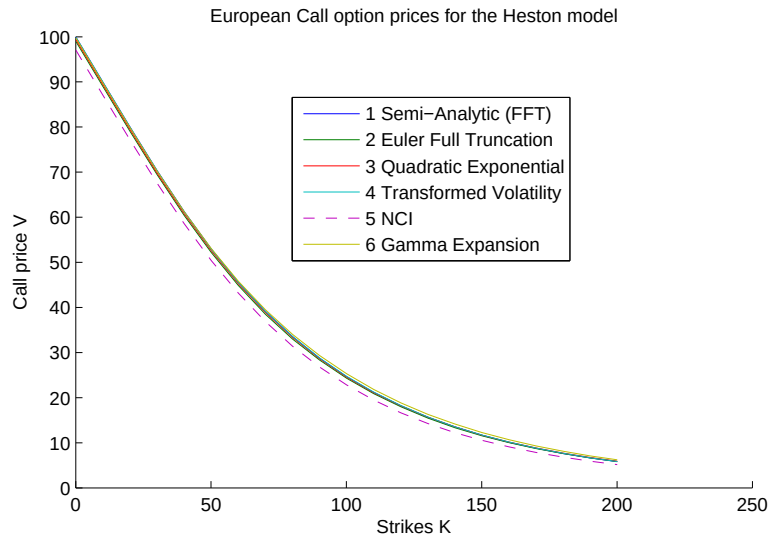


Figure 3.1: European call price estimates, obtained using different schemes

The estimates were all obtained by performing a Monte Carlo run, where we used 10.000 paths, $\Delta = 1$ and the following parameter set:

Parameter	$S(0)$	$r(t)$	κ	$\theta = V(0)$	ε	ρ
Value	100	0	0.5	0.04	0.1	-0.1

Table 3.1: Parameter values of the underlying of the European call option

The maturity of the option is taken to be 1 year.

Remark. Note that from the parameter set in Figure 3.1, the NCI option price differs significantly from the analytic price: we found that the scheme is biased because we use an equidistant grid. For small realizations of $V(t + \Delta)$, the interpolation grid of u_{N_j} should be refined near zero to obtain sufficient accuracy. The NCI scheme did perform quite well for other parameter sets. However, since NCI will not be our candidate scheme for the multi-dimensional case, we did not implement this improvement.

We tested the performance of the schemes for cases for which the Feller condition was satisfied and for which it was violated. The accuracy and CPU time of the pricing of a European call option was measured for all schemes. The reference price is the semi-analytic price obtained by the Carr-Madan method (see [Carr & Madan, 1999]). The results are displayed (qualitatively) in Table 3.2.

	FT	TV	QE	NCI	GE
European call price accuracy (Feller satisfied)	+	+	+	-	-
European call price accuracy (Feller violated)	-	-	+	+	+
CPU time	+	+	+	-	-
Easily applicable for multi-D	+	+	?	?	-

Table 3.2: Qualitative comparison of the considered Monte Carlo methods.

There is a significant difference in accuracy for situations in which the Feller condition is and in which it is not satisfied. Therefore, we tested the schemes for at least the following two cases: one with small volatility of volatility ($\varepsilon = 0.1$), and one with high volatility of volatility ($\varepsilon = 1$). In both cases, $S(0) = 100$, $\kappa = 0.5$, $\theta = V(0) = 0.04$, $T = 10$, $\Delta = 1/16$ and the number of paths is 10.000.

The Euler FT scheme is inaccurate when the Feller condition is violated. The TV scheme tends to blow up the variance process whenever $4\kappa\theta < \varepsilon^2$. This is a serious drawback of the TV scheme. On the other hand, NCI and GE do not perform very well when ε is too small with respect to $\kappa\theta$, since then the schemes slow down so much that a memory limit breaks down the scheme.

One generic method would be preferable over switching between more than one method. In the latter case namely, the option value, as a function of ε , may show a discontinuity at the value of ε where the scheme switches from one method to another. This discontinuity is unrealistic and therefore undesirable. The QE scheme has the advantage of being robust in this sense.

In terms of CPU time, Euler FT, TV and QE always perform well because of their relative low computational effort per time step. NCI and GE are often reasonably fast, but again if ε is small with respect to $\kappa\theta$, the computation can become very slow.

One other comparison should be made: for a multi-dimensional extension (i.e., vector-valued SDEs), all processes should be correlated correctly. For Euler FT and TV, this can of course be done by the use of a Cholesky decomposition. The same cannot be done in case the variance scheme is handled in the QE, NCI or GE schemes, since then not all random variables are normally distributed.

Instead, one could use the normal copula to correlate these random variables, but this method is non-exact. The resulting correlation differs from the imposed correlation. This difference is known to be small, which is why the problem may be surmountable. Therefore, QE and NCI have been given a question mark for this feature. GE has got a minus, since its use of many random variables implies the need of an extensive research of how to correlate the variance processes appropriately.

We could also have distinguished biases, which formally denote the difference between the expectation of an estimator and the parameter which is estimated. Loosely speaking, this means that a scheme contains no bias when any estimation error can be reduced by an increase of the number of paths.

Each method contains bias in some sense: even the numerical estimate of the exact representation in [Broadie & Kaya, 2006] (see Section 3.2.1) has an expectation differing from the correct value, due to the truncation and trapezoidal rule in (3.6).

Euler FT is of course subject to bias whenever variance paths contain negative values. TV has bias since the model which it is based on, does not correspond to the Heston model. The QE scheme approximates the variance distribution and the integral over the variance path, NCI interpolates and GE truncates infinite sums. It is hard to extract some performance measure from these different biases, which is why we do not take this issue into account.

Mainly based on the robustness in accuracy and speed, we choose QE as the candidate scheme for our multi-stock Heston Monte Carlo method.

Chapter 4

The MQE Monte Carlo method

Consider D stocks, each following the Heston dynamics as defined in Expressions (2.1) and (2.2). We already pointed at the problem of the correlations between the processes of a D -dimensional Heston model. In fact, the very difference between a D -dimensional Heston model and D one-dimensional Heston models lies in the correlation structure of the model. We can distinguish:

1. ρ_i , the correlations between the Brownian motions of the i -th stock and variance, just as we had in the one-dimensional Heston model;
2. ρ_{ij} , the correlations between the i -th and j -th stock Brownian motions;
3. $\rho_{V_i V_j}$, the correlations between the i -th and j -th variance Brownian motions;
4. $\rho_{S_i V_j}$, the correlations between the Brownian motions of the i -th stock and the j -th variance, $i \neq j$.

When we consider D one-dimensional Heston models, we assume that the latter three correlations are zero. When constructing a multi-dimensional Heston Monte Carlo method, we assume that at least some of these correlations can be non-zero. First, we will discuss the significance of these different correlations.

4.1 Assumptions

Just as in the case of one stock, correlations between a stock and its variance, ρ_i , are also significant in case of multiple stocks. We explained in Section 2.2 how we can ensure that the correlation estimate of our Monte Carlo paths will not structurally differ from the imposed correlation. We have to rewrite the stock dynamics so that the input Brownian motion is independent of the other random variables. When we use the QE method in particular, we can illustrate this as follows. Instead of writing

$$X_i(t + \Delta) = X_i(t) + \int_t^{t+\Delta} \left(r(s) - \frac{1}{2} V_i(s) \right) ds + \int_t^{t+\Delta} \sqrt{V_i(s)} dW_{X_i}(s), \quad (4.1)$$

with $dW_{X_i}(t)dW_{V_i}(t) = \rho_i dt$, we could equally write

$$\begin{aligned} X_i(t + \Delta) = & X_i(t) + \int_t^{t+\Delta} r(s) ds + \frac{\rho_i}{\epsilon_i} (V_i(t + \Delta) - V_i(t) - \kappa_i \theta_i \Delta) \\ & + \left(\frac{\kappa_i \rho_i}{\epsilon_i} - \frac{1}{2} \right) \int_t^{t+\Delta} V_i(s) ds + \sqrt{1 - \rho_i^2} \int_t^{t+\Delta} \sqrt{V_i(s)} dW_i(s), \end{aligned} \quad (4.2)$$

with $W_i(t)$ an independent Brownian motion. Therefore, both expressions (4.1) and (4.2) can be discretized to sample the next step stock value. However, suppose we would use the QE method and discretize (4.1), thus obtaining

$$X_i(t + \Delta) = X_i(t) + \left(r(t) - \frac{1}{2} V_i(t) \right) \Delta + \sqrt{V_i(t)} \sqrt{\Delta} Z_{X_i}.$$

Then we would correlate the standard normal random variables Z_{X_i} and Z_{V_i} by setting $Z_{X_i} = \rho_i Z_{V_i} + \sqrt{1 - \rho_i^2} Z_i$, with Z_i a standard normal random variable, which is independent of Z_{V_i} . However, since we apply a non-linear transformation with Z_{V_i} in the QE method (by setting $\hat{V}_i(t + \Delta) = a(b + Z_{V_i})^2$, for example), the resulting correlation will differ from the imposed correlation.

This phenomena has been named *leaking correlation*, and it is avoided in [Broadie & Kaya, 2006] by discretizing representation (4.2), which has independent random variables as input. We will show how we can extend this procedure to the multi-dimensional case in the next section.

Correlations between stocks can certainly not be neglected, for the obvious reason that different markets influence each other. Consider a foreign exchange option for example: one exchange rate will influence another exchange rate. If one specific part of the market goes up or down, one often observes that another particular stock follows this movement to some extent, or shows the opposite behavior. Let us discuss how to implement the stock-stock correlation in our multi-dimensional Monte Carlo method. One can see in (4.2) that when $X_i(t), V_i(t), V_i(t + \Delta)$ and $\int_t^{t+\Delta} V_i(s) ds$ are given, the next log-stock value $X_i(t + \Delta)$ is normally distributed: the first four terms on the right-hand side are then known, while the remaining term is a constant times an Itô integral with a deterministic integrand. This term has thus a normal distribution. To correlate these normal random variables appropriately, we can use a Cholesky decomposition of Σ_X , the correlation matrix of the stock Brownian motions. We will discuss this idea in detail in the next section.

Correlations between variances may also be significant: when a basket of stocks becomes more volatile on average, some traders may react by trading other stocks as well, which in turn will increase their volatility. The implementation of correlated variance processes in our Monte Carlo method will not be straightforward. With the QE method, the corresponding random variables are not normally distributed. Therefore, an approach based on a Cholesky decomposition will not impose the desired correlation exactly. Nevertheless, a well-known procedure to correlate random variables which are not normally distributed, is the NorTA method, which employs the normal copula:

1. Generate a standard normal random vector $\mathbf{Z} \in \mathbb{R}^D$.
2. Perform a Cholesky decomposition on $\Sigma_V = LL^T$, the correlation matrix of the variance Brownian motions. Set $\bar{\mathbf{Z}} = L\mathbf{Z}$, then $\bar{\mathbf{Z}}$ is a D -dimensional standard normal vector with a correlation structure defined by Σ_V .
3. Transform $\bar{\mathbf{Z}}$ to $\mathbf{X}$, a random vector with the desired marginal distributions, by computing element-wise $X_i := F_i^{-1}(\Phi(\bar{Z}_i))$ for $i = 1, \dots, D$, with $\Phi(x)$ the standard normal cumulative distribution function and F_i the cumulative distribution function of the i -th desired marginal distribution.

See [Ghosh & Henderson, 2003] or [Chen, 2001], for example. The correlation between X_i and X_j will differ from the correlation between $\bar{Z}_i$ and $\bar{Z}_j$. The difference is expected to be small, though we do not know its exact size. We could use the normal copula method and investigate this difference. In the MQE scheme however, we set the variance-variance correlations equal to zero. In this way, our investigation of the performance of the MQE scheme will not be influenced by an error that this NorTA method would induce.

Finally, we assume that correlations between stocks and other variances will be zero as well.

4.2 Construction of the MQE scheme

By extending the one-dimensional QE method in [Andersen, 2007] to the D -dimensional case, we aim at a sampling scheme in which we go through 3 stages each time step and path:

1. Firstly, we sample the next step of each variance process independently, starting with $V_1(t + \Delta)$, followed by $V_2(t + \Delta)$, and so on.
2. Secondly, we estimate all variance path integrals during this time step, $\int_t^{t+\Delta} V_i(s)ds$, $i = 1, \dots, D$.
3. Third of all, we sample the next step of each stock process, properly correlated with its corresponding variance process as well as with all other stock processes. First $S_1(t + \Delta)$, then $S_2(t + \Delta)$, and so on.

The order of these stages will be of crucial importance in our reasoning.

4.2.1 Set up of the MQE scheme

For $i = 1, \dots, D$, the dynamics of the i -th stock are given by

$$dV_i(t) = \kappa_i(\theta_i - V_i(t))dt + \varepsilon_i \sqrt{V_i(t)} dW_{V_i}(t), \quad (4.3)$$

$$dX_i(t) = \left(r(t) - \frac{1}{2}V_i(t)\right)dt + \sqrt{V_i(t)} dW_{X_i}(t), \quad (4.4)$$

and $\Sigma \in \mathbb{R}^{2D \times 2D}$ is the correlation matrix of the $2D$ -dimensional Brownian motions vector

$$\begin{pmatrix} dW_{V_1}(t) \\ \vdots \\ dW_{V_D}(t) \\ dW_{X_1}(t) \\ \vdots \\ dW_{X_D}(t) \end{pmatrix}.$$

The mentioned correlation assumptions imply that Σ will be of the form

$$\Sigma = \begin{pmatrix} I_D & \Sigma_{XV} \\ \Sigma_{XV}^\top & \Sigma_X \end{pmatrix}, \quad (4.5)$$

with I_D the $D \times D$ -identity matrix, Σ_X the $D \times D$ -correlation matrix of the stocks, and Σ_{XV} a diagonal matrix of the same size, given by

$$\Sigma_{XV} = \begin{pmatrix} \rho_1 & 0 & \dots & 0 \\ 0 & \rho_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \rho_D \end{pmatrix}.$$

A Cholesky decomposition $\Sigma = LL^\top$ can be used to correlate independent Brownian motions appropriately:

$$\begin{pmatrix} dW_{V_1}(t) \\ \vdots \\ dW_{V_D}(t) \\ dW_{X_1}(t) \\ \vdots \\ dW_{X_D}(t) \end{pmatrix} \stackrel{d}{=} L d\tilde{W}(t) = L \begin{pmatrix} d\tilde{W}_1(t) \\ \vdots \\ d\tilde{W}_{2D}(t) \end{pmatrix}, \quad (4.6)$$

with $d\tilde{W}(t)$ a vector of independent Brownian motion increments. L is a $2D \times 2D$ lower triangular matrix. We will now state a proposition about the structure of L .

Proposition 2. Let us assume a correlation structure as stated at the start of this section. Consider a Cholesky decomposition $LL^\top$ of the corresponding correlation matrix Σ from (4.5). Then L is a $2D \times 2D$ lower triangular matrix which can be written as the following block matrix:

$$L = \begin{pmatrix} I_D & O_D \\ \Sigma_{XV} & L^* \end{pmatrix}, \quad (4.7)$$

with O_D the $D \times D$ zero matrix and L^* a $D \times D$ lower triangular matrix.

Proof. In Appendix A, we have proved that for a correlation matrix, there is a lower triangular matrix L so that $\Sigma = LL^\top$. We write out the equation $\Sigma = LL^\top$ in an element-wise fashion, thus obtaining the following formulas¹ for the elements of L :

$$L_{j,j} = \sqrt{\Sigma_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2}, \quad L_{i,j} = \frac{1}{L_{j,j}} \left(\Sigma_{i,j} - \sum_{k=1}^{j-1} L_{i,k} L_{j,k} \right), \quad \text{for } i > j, \quad j = 1, \dots, 2D.$$

The elements of the two upper submatrices in (4.7) follow immediately from these two formulas. For the elements of the lower left submatrix of L , we can simplify the second formula above to

$$L_{D+i,j} = \frac{1}{L_{j,j}} \left(\Sigma_{D+i,j} - \sum_{k=1}^{j-1} L_{D+i,k} L_{j,k} \right), \quad \text{for } D+i > j, \quad j = 1, \dots, D. \quad (4.8)$$

Since the upper left submatrix of L is I_D , it follows that $L_{j,j} = 1$ and $L_{j,k} = 0$ for $j = 1, \dots, D, k = 1, \dots, j-1$. Substitution in (4.8) yields that $L_{D+i,j} = \Sigma_{D+i,j}$ for $i = D+1, \dots, 2D, j = 1, \dots, D$, which is what we had to show for the lower left submatrix. The lower right submatrix is lower triangular since L is lower triangular, which was the only constraint that we claimed for this submatrix. $\square$

The elements of L depend on the correlations only. For example, if $D = 2$, then

$$\Sigma = \begin{pmatrix} 1 & 0 & \rho_1 & 0 \\ 0 & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & \rho_{12} \\ 0 & \rho_2 & \rho_{12} & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \rho_1 & 0 & \sqrt{1-\rho_1^2} & 0 \\ 0 & \rho_2 & \frac{\rho_{12}}{\sqrt{1-\rho_1^2}} & \sqrt{\frac{(1-\rho_1^2)(1-\rho_2^2)-\rho_{12}^2}{1-\rho_1^2}} \end{pmatrix}.$$

4.2.2 Simulation of the next time step

The first stage of the MQE scheme is the sampling of the next step variance, $V_i(t+\Delta)$. Note that the upper-left identity submatrix of L in (4.7) confirms the fact that the variance processes can be simulated independently. In the MQE scheme, we use the QE method D times every step with independent random variables as input.

The second stage of the MQE scheme is the simulation of the variance path integral $\int_t^{t+\Delta} V_i(s) ds$. Since an exact simulation is computationally expensive (see [Broadie & Kaya, 2006]). In [Glasserman & Kim, 2008], the author proposes an almost exact method to sample this quantity. As described in Section 3.2.4, this method is not robust. Instead, we will use the intuitive (central discretization) approximation in [Andersen, 2007]:

$$\int_t^{t+\Delta} \hat{V}_i(s) ds \approx \Delta[\gamma_1 \hat{V}_i(t) + \gamma_2 V_i(t+\Delta)], \quad (4.9)$$

with $\gamma_1 = \gamma_2 = \frac{1}{2}$.

¹These formulas are also known from the Cholesky-Banachiewicz algorithm, which is another algorithm to perform a Cholesky decomposition than the one we describe in Appendix A.

In the final stage, we focus on the simulation of the next step of the stocks. We would like to sample $X_1(t + \Delta)$ first, then $X_2(t + \Delta)$, and so on, with the independent Brownian motion vector $d\tilde{W}(t)$ as input. Therefore, we first substitute (4.7) in (4.6), to obtain

$$\begin{pmatrix} dW_{V_1}(t) \\ \vdots \\ dW_{V_D}(t) \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} d\tilde{W}_1(t) \\ \vdots \\ d\tilde{W}_D(t) \end{pmatrix}, \quad (4.10)$$

$$\begin{pmatrix} dW_{X_1}(t) \\ \vdots \\ dW_{X_D}(t) \end{pmatrix} \stackrel{d}{=} \Sigma_{XV} \begin{pmatrix} d\tilde{W}_1(t) \\ \vdots \\ d\tilde{W}_D(t) \end{pmatrix} + L^* \begin{pmatrix} d\tilde{W}_{D+1}(t) \\ \vdots \\ d\tilde{W}_{2D}(t) \end{pmatrix}. \quad (4.11)$$

One can see from (4.10) that the Cholesky decomposition assures that the variance processes are driven by independent Brownian motions. We write out the i -th row of (4.11), obtaining

$$dW_{X_i}(t) \stackrel{d}{=} \rho_i d\tilde{W}_i(t) + \sum_{j=1}^i L_{i,j}^* d\tilde{W}_{D+j}(t),$$

where we used the facts that Σ_{XV} is a diagonal matrix and L^* a lower triangular matrix. Substitution in (4.4), yields

$$dX_i(t) = \left(r(t) - \frac{1}{2} V_i(t) \right) dt + \sqrt{V_i(t)} \rho_i d\tilde{W}_i(t) + \sum_{j=1}^i \sqrt{V_i(t)} L_{i,j}^* d\tilde{W}_{D+j}(t).$$

The integrated form reads

$$X_i(t + \Delta) = X_i(t) + \int_t^{t+\Delta} r(s) - \frac{1}{2} V_i(s) ds + \rho_i \int_t^{t+\Delta} \sqrt{V_i(s)} d\tilde{W}_i(s) + \sum_{j=1}^i L_{ij}^* \int_t^{t+\Delta} \sqrt{V_i(s)} d\tilde{W}_{D+j}(s). \quad (4.12)$$

Note that we would like to sample $X_i(t + \Delta)$ after the realization of $\mathcal{V}_i := \{V_i(t), V_i(t + \Delta), \int_t^{t+\Delta} V_i(s) ds\}$, which is done in the first two stages of the MQE scheme. Therefore, to sample the left-hand side of (4.12) exactly given $\mathcal{V}_i$, one has to sample the Itô integrals on the right-hand side exactly given $\mathcal{V}_i$. We can distinguish two types of Itô integrals: those with the Brownian motion of a variance process and those with the Brownian motion of a stock process. Let us name them *variance Itô integrals* and *stock Itô integrals*, respectively.

Variance Itô integrals

After having simulated the next step variance with the QE method, the Brownian motion of the variance process has not been sampled explicitly, but the substitute input random variable of the QE method *has*. Therefore, we learned from the one-dimensional case in [Andersen, 2007], not to sample this Brownian motion directly in this stage, since then correlation may leak. We also learned (see formula (2.7)) that by integrating the variance process, we obtain an exact expression for each variance Itô integral:

$$\begin{aligned} \int_t^{t+\Delta} \sqrt{V_i(s)} d\tilde{W}_i(s) &= \int_t^{t+\Delta} \sqrt{V_i(s)} dW_{V_i}(s) \\ &= \varepsilon_i^{-1} \left(V_i(t + \Delta) - V_i(t) - \kappa_i \theta_i \Delta + \kappa_i \int_t^{t+\Delta} V_i(s) ds \right) \quad \forall i = 1, \dots, D, \end{aligned} \quad (4.13)$$

where the first equality holds since all variance processes are independent. In this way, we can compute the exact value of the variance Itô integrals given $\mathcal{V}_1, \dots, \mathcal{V}_D$.

Stock Itô integrals

All other Itô integrals on the right-hand side of (4.12) are stock Itô integrals, and they are all contained under the summation term. In the one-dimensional case, only the final term under this summation in (4.12) is present, which is an independent normal random variable. However, when propagating more than one stock, we consider groups of

different stock Itô integrals with identical Brownian motions. To simulate one group of these stock Itô integrals, we must use the same normal random variable sample for each integral, and multiply it by different constants which are chosen so that the variances correspond with those of these different stock Itô integrals. Formally, given $\mathcal{V}_i$,

$$\int_t^{t+\Delta} \sqrt{V_i(s)} d\tilde{W}_{D+j}(s) \stackrel{d}{=} Z_j \sqrt{\int_t^{t+\Delta} V_i(s) ds}, \quad \forall i, j = 1, \dots, D, \quad (4.14)$$

with $\mathbf{Z} = (Z_1 \dots Z_D)^\top$ an independent standard normal random vector. By sampling this random vector exactly, we can sample all stock Itô integrals, given $\mathcal{V}_1, \dots, \mathcal{V}_D$, exactly.

Matrix-vector form

Numerical implementation is efficient in vector form. To end up with this form, suppose that $\mathcal{V}_1, \dots, \mathcal{V}_D$ are sampled, and let us define, for $i = 1, \dots, D, j = 1, \dots, D$,

$$\begin{aligned} I_{i,j} &:= \int_t^{t+\Delta} V_i(s) d\tilde{W}_j(s), \\ s_i &:= \sqrt{\int_t^{t+\Delta} V_i(s) ds}, \\ f_i &:= \epsilon_i^{-1} \left(V_i(t+\Delta) - V_i(t) - \kappa_i \theta_i \Delta + \kappa_i \int_t^{t+\Delta} V_i(s) ds \right), \\ Z_j &\sim \mathcal{N}(0, 1). \end{aligned}$$

For all i , given $\mathcal{V}_i$, s_i represents the standard deviation of the Itô integral $I_{i,j}$. Then Expressions (4.13) and (4.14) can, respectively, be written as

$$\begin{aligned} I_{i,i} &= f_i & \text{for } i = 1, \dots, D, \\ I_{i,D+j} &= s_i Z_j & \text{for } i, j = 1, \dots, D. \end{aligned}$$

After substituting this in (4.12), the last two terms of the right-hand side in (4.12) read

$$\rho_i I_{i,i} + \sum_{j=1}^i L_{ij}^* I_{i,D+j} = \rho_i f_i + \sum_{j=1}^i L_{ij}^* s_i Z_j \quad \forall i = 1, \dots, D.$$

Therefore, (4.12) can be written, in vector notation, as

$$\mathbf{X}(t+\Delta) = \mathbf{X}(t) + \int_t^{t+\Delta} \mathbf{r}(s) ds - \frac{1}{2} \mathbf{V}(s) ds + \boldsymbol{\rho}^\top \mathbf{f} + D_s L^* \mathbf{Z}, \quad (4.15)$$

with

$$\mathbf{X}(t) = \begin{pmatrix} X_1(t) \\ \vdots \\ X_D(t) \end{pmatrix}, \mathbf{V}(t) = \begin{pmatrix} V_1(t) \\ \vdots \\ V_D(t) \end{pmatrix}, \boldsymbol{\rho} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_D \end{pmatrix}, \mathbf{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_D \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_D \end{pmatrix},$$

and

$$D_s = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & s_D \end{pmatrix}.$$

If $\mathcal{V}_1, \dots, \mathcal{V}_D$ and $\mathbf{X}(t)$ are given, then the first three terms on the right-hand side of (4.15) are known while the fourth term can be simulated exactly by drawing an independent standard normal random vector $\mathbf{Z}$. An element-wise implementation of the third stage of the MQE scheme is then based on (4.12), whereas a vector-wise implementation will

employ (4.15).

The MQE simulation of the next step variances and the variance path integrals goes completely analogously to the corresponding simulation schemes of the conventional QE method. The remaining part of the MQE scheme is the propagation of all stock paths. In this section, we constructed a method to simulate these next steps of the log-stock values exactly, given $\bigcup_{i=1}^D \mathcal{V}_i$. We recall that the distribution of the next step log-stock value of stock i is known just when $\mathcal{V}_i$ is given. This explains the importance of the sampling order that we claimed at the start of this section.

Chapter 5

Performance tests for the MQE scheme

Let us recall the approximations on which the MQE scheme is based. Since there are no additional approximations required in the multi-dimensional extension of the stock propagation, the only approximations are those in the first two stages. These approximations are all immediately inherited from the QE method, which have shown to perform satisfactory in [Andersen, 2007]. Furthermore, as is clear from the algorithm, the number of computations is linear in the model dimension. This means that the CPU time will not grow exponentially as the dimensionality grows. Based on these arguments, we have confidence that the MQE scheme may perform well. Nevertheless, we will test our method extensively in this chapter.

5.1 Imposed correlation test

We implement a two-dimensional QE method and check whether the estimated correlations behave as expected. The estimated correlations obtained by using a two-dimensional Euler FT scheme serve as the reference values. We consider the following two cases:

	<i>Feller Satisfied</i>	<i>Feller Violated</i>
$S_i(0)$	100	100
$r(t)$	0	0
κ_i	0.5	0.5
$\theta_i, V_i(0)$	0.04	0.04
ε_i	0.01	1

Table 5.1: Parameter settings of both stocks S_1 and S_2

One can see that the parameter set *Feller Violated* is the multi-dimensional version of Test Case 1 in [Andersen, 2007]. Parameter set *Feller Satisfied* is as *Feller Violated*, but now the volatility of volatility is set such that the Feller condition is satisfied. In both cases, we set $T = 10$ (years), the number of Monte Carlo paths is 10.000, $\Delta_{QE} = 46$ days (2^{-3} year), $\Delta_{FT} = 6$ days (2^{-6} year) and the correlation matrix is given by

$$\Sigma = \begin{pmatrix} 1 & \rho_{V_1 V_2} & \rho_1 & \rho_{V_1 S_2} \\ \rho_{V_1 V_2} & 1 & \rho_{V_2 S_1} & \rho_2 \\ \rho_1 & \rho_{V_2 S_1} & 1 & \rho_{12} \\ \rho_{V_1 S_2} & \rho_2 & \rho_{12} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -0.3 & 0 \\ 0 & 1 & 0 & -0.6 \\ -0.3 & 0 & 1 & 0.7 \\ 0 & -0.6 & 0.7 & 1 \end{pmatrix}.$$

We will use these parameter sets more often in this thesis and refer to them by the names *Feller Satisfied* and *Feller Violated*, respectively.

5.1.1 Results of imposed correlation test

We estimate the correlation between the SDEs of the two stocks and the two correlations between the stock SDE and corresponding variance SDE by using numerical schemes. The computation of correlation estimates is based on the definition of correlation:

$$\text{corr}(Y_i, Y_j) := \frac{\text{Cov}(Y_i, Y_j)}{\sqrt{\text{Cov}(Y_i, Y_i)\text{Cov}(Y_j, Y_j)}}, \text{ with } \text{Cov}(Y_i, Y_j) = \mathbb{E}[(Y_i - \mathbb{E}[Y_i])(Y_j - \mathbb{E}[Y_j])].$$

The estimated correlations of *Feller Satisfied* are shown in Figure 5.1.

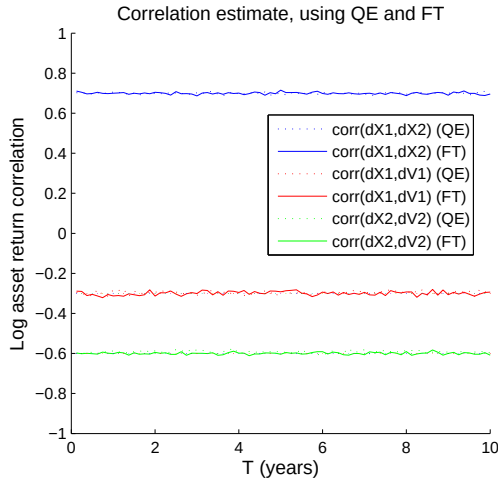


Figure 5.1: Correlation estimates in *Feller Satisfied*, using QE and Euler FT

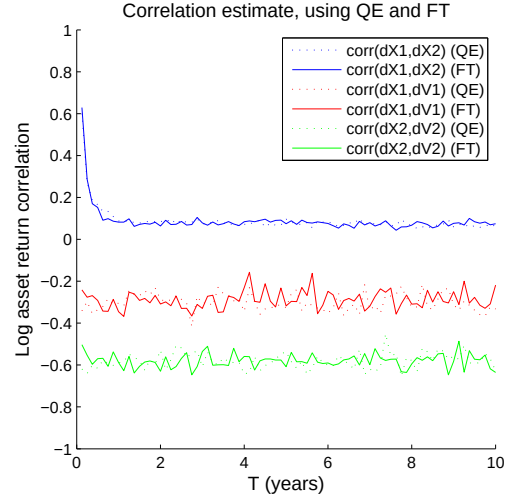


Figure 5.2: Correlation estimates in *Feller Violated*, using QE and Euler FT

As one can see in this figure, all correlation estimates of the MQE and Euler FT scheme are approximately equal, which is as expected. The same holds for *Feller Violated* in Figure 5.2, but now $\text{corr}(dX_1(t), dX_2(t))$ drops from an initial value of approximately ρ_{12} towards zero during the first couple of time steps. Although this may look strange at first sight, it is expected. To understand this, first note that the $\text{corr}(dX_1(t), dX_2(t))$ denotes the log-stock return correlation. Assuming Black-Scholes dynamics, this correlation is equal to the correlation between the corresponding Brownian motions. Since the variance term in each SDE is stochastic, we have

$$\begin{aligned} \text{corr}(dX_1(t), dX_2(t)) &= \text{corr}((r - \sigma_1^2)dt + \sigma dW_{X_1}(t), (r - \sigma_2^2)dt + \sigma_2 dW_{X_2}(t)) \\ &= \text{corr}(dW_{X_1}(t), dW_{X_2}(t)) \\ &= \mathbb{E}[dW_{X_1}(t)dW_{X_2}(t)]/dt \\ &= \rho_{12}. \end{aligned} \tag{5.1}$$

In the second equality, we used the fact that r, σ_1, σ_2 are deterministic. However, for Heston dynamics, the first equality does not hold, since then the volatility is stochastic. Therefore under the Heston model, the log-stock return correlation is theoretically different from the correlation between the two corresponding Brownian motions. In *Feller Violated*, the simulated variance path can often reach zero, which drives the log-stock return correlation towards zero, regardless of the imposed correlation value on the stock Brownian motions. Instead, when the Feller condition is satisfied, this stochastic variance cannot be absorbed in the origin, while its mean-reverting property makes it reverting to the long-term variance. Therefore, the correlation between the SDEs more or less coincides with ρ_{12} in that case. This explains the difference between the two blue, most upper function graphs in Figure 5.1 and those in Figure 5.2.

Despite the confusion, one might say that this correlation drop is not troublesome as long as the model prices fit market prices. However, suppose we would like to investigate the sensitivity of an option value to different models. Then to

compare a multi-asset Heston value fairly with, say, a multi-asset Black-Scholes value, it is hard to make sure that the log-stock return correlations mutually coincide.

Now one may wonder why the estimated correlations between log-stock return and corresponding variance return do not decrease in time. The explanation is that the variance SDE contains *the same* stochastic variance term as the stock SDE. Therefore, no matter how close the variance is to zero, both Brownian motions $dW_{S_i}(t)$ and $dW_{V_i}(t)$ are scaled by the same factor. This does not cause any reduction in correlation. In fact, one can check that the correlation estimate between the two stock processes does not decrease if we take $V_1(t) = V_2(t)$.

At last, note that the estimates for $\text{corr}(dX_1(t), dV_1(t))$ and $\text{corr}(dX_2(t), dV_2(t))$ vary in time in *Feller Violated*. One can explain this fact by the more volatile variance paths in *Feller Violated*. The variance path exhibits significant jumps every time step, which increases the standard error of the correlation estimate.

Just like the Euler FT scheme, the MQE scheme shows the expected correlation estimates. Therefore, we conclude that the MQE scheme has passed this test.

5.2 Option validation: the *Feller Satisfied* Test

In a one-dimensional Heston model, one can check the option pricing performance of a Monte Carlo method by comparing the results with those obtained by the Carr-Madan method ([Carr & Madan, 1999]) or the COS method ([Fang & Oosterlee, 2008]). A European call or put option can be priced by one of these semi-analytic Fourier methods, and this price can serve as a reference value for the Monte Carlo result.

With multi-dimensional Heston dynamics however, the characteristic function of the model (which is necessary in any Fourier method) is not known in general, which makes this procedure inapplicable. As an alternative, we try to obtain a reference by use of the Euler FT scheme with a very small time step and a large number of Monte Carlo paths, as we expect that the bias will be reduced in this way. However, for high volatility of volatility parameter (even for very fine time grids and many paths), the Euler FT variance path will often become negative, as in the one-dimensional case. This will cause bias, which is why we prefer not to use Euler FT as a reference method when the Feller condition is not fulfilled.

For our first test, we choose the parameter values of *Feller Satisfied*, and use the Euler FT scheme as a reference for the MQE scheme. The payoff function that we will evaluate, is given by

$$\max[S_1(T) + S_2(T) - K, 0], \quad (5.2)$$

for some strike price K at maturity T . Results are presented in Figure 5.3.

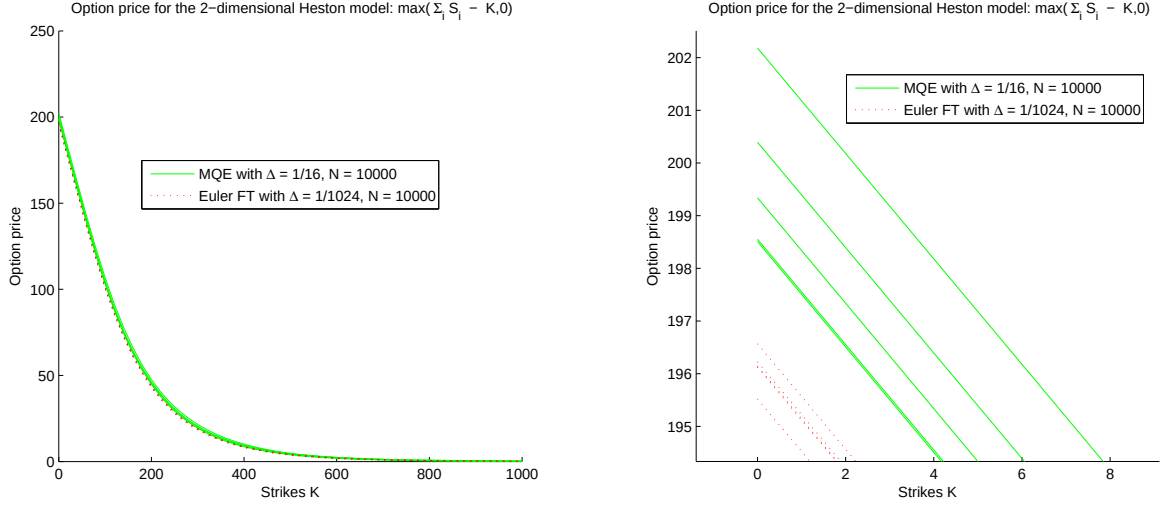


Figure 5.3: Basket option price: $\rho_{12} = 0.7, S_i(0) = 100, \kappa_i = 1.5, \theta_i = V_i(0) = 0.04, \epsilon_i = 0.01, \rho_i = -0.3$ for $i = 1, 2$. The right-hand figure zooms in on the upper left corner of the left-hand figure.

In both plots of Figure 5.3, five MQE option prices are plotted against five Euler FT option prices. One can see from the left-hand plot that both estimates coincide quite well. The Euler FT prices are slightly lower than the MQE prices. However, even this small difference may be caused by a bias of the Euler FT scheme. The right-hand plot in Figure 5.3 shows that the MQE prices are closer to the analytic value of the payoff function at $K = 0$ (which is 200) than the Euler FT prices. This may imply that the Euler FT prices have some negative bias, since all estimates are smaller than 200 at $K = 0$.

We conclude that the MQE scheme has successfully passed an option validation test in case *Feller Satisfied*.

5.3 Option validation: the *Feller Violated* Test I

Since we already stated that the Feller condition is often violated in financial practice, we would also like to test the MQE scheme when the Feller condition is violated. Furthermore, we stated in the previous section that we cannot use the Euler FT scheme as a reference for the MQE scheme, due to its large bias when the Feller condition is violated. Therefore, we perform two other tests. The first test will be treated in this section.

Consider a European call option on the sum of several underlyings. The payoff function is then:

$$\max \left[\sum_{i=1}^D S_i(T) - K, 0 \right].$$

Suppose we make the (somewhat trivial) assumption that all stocks are identical. Then the problem becomes one-dimensional immediately, and we can thus find the semi-analytic price (by use of a Fourier method). In that case, we can find a reference value for the MQE prices. Of course, no basket option will be of this form in practice, but this example nevertheless serves as a first test for the MQE scheme under *Feller Violated*.

We choose $D = 3, T = 10$ years and the following parameter values per stock:

Parameter	$S_i(0)$	$r(t)$	κ_i	$\theta_i, V_i(0)$	ϵ_i
Value	100	0	0.5	0.04	1

Table 5.2: Parameter values of all stocks S_1, S_2 and S_3

These values coincide with *Feller Violated*, but now the correlation structure is different. In order to have identical stocks, one has to choose

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 & \rho_1 & \rho_1 & \rho_1 \\ 1 & 1 & 1 & \rho_1 & \rho_1 & \rho_1 \\ 1 & 1 & 1 & \rho_1 & \rho_1 & \rho_1 \\ \rho_1 & \rho_1 & \rho_1 & 1 & 1 & 1 \\ \rho_1 & \rho_1 & \rho_1 & 1 & 1 & 1 \\ \rho_1 & \rho_1 & \rho_1 & 1 & 1 & 1 \end{pmatrix}.$$

One may be confused since a full correlation matrix is not allowed in the MQE method. However, the correlations can be attained in this case, by using the same Brownian motions for each variance process. We choose $\rho_1 = -0.6$. The results are shown in Figure 5.4.

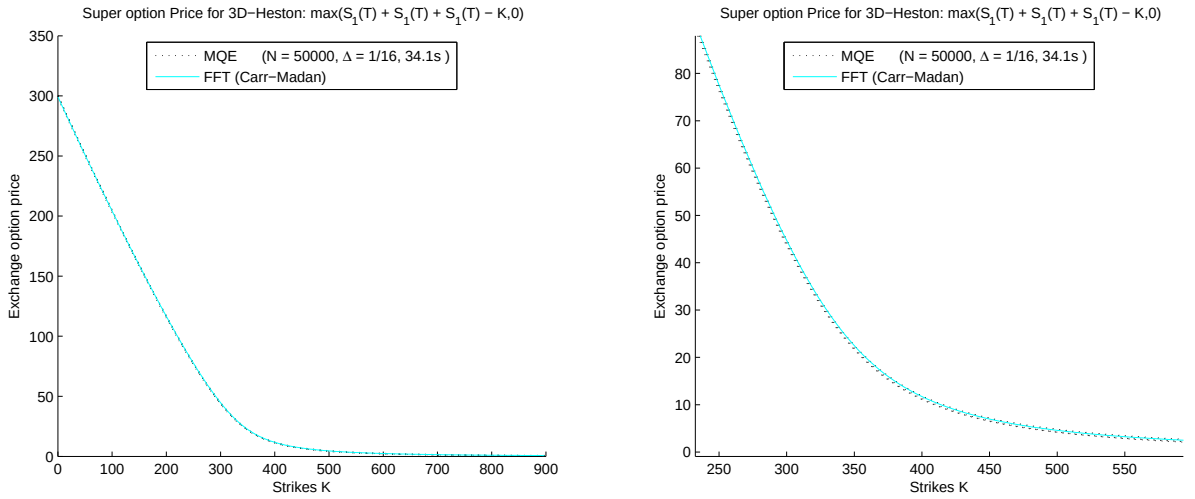


Figure 5.4: Basket option price estimates when Feller condition is violated, all three stocks identical. The right-hand plot zooms in on the left-hand plot.

One can see in the left-hand plot of this figure that for $N = 50.000$ and $\Delta = 1/16$, the MQE option price practically coincides with the semi-analytic option price. In fact, the semi-analytic price is contained in the 95% confidence interval of the price as estimated from the Monte Carlo Paths. This shows the absence of a significant error statistically. Since the confidence interval is relatively small (less than 3% of the semi-analytic price for in-the-money option, and less than 30% of the semi-analytic price for out the money options), we conclude that the MQE scheme has passed this test successfully.

5.4 Option validation: the *Feller Violated* Test II

The test of the previous section is not very general, since the three-dimensional problem is actually one-dimensional. In this section, we perform a second test: we will reduce a two asset Heston model to a single-asset model, by changing the numéraire. In this way, we can find the basket option price semi-analytically, thus obtaining a reference value for the MQE estimates for option prices with two correlated stocks as underlyings. We will see that we have to make some restrictions on the two-dimensional Heston parameter set. However, since we will be allowed to choose any value for the correlation between the two stocks, ρ_{12} , this test will be more general than that of Section 5.3.

5.4.1 Problem set up

Suppose that $B(t)$ is the price at time t of one unit of the currency that was invested in the money market at time $t = 0$. Then

$$dB(t) = r(t)B(t)dt,$$

and thus $B(t) = e^{\int_0^t r(s)ds}$. It is common to define $B(t)$ as the numéraire¹ under the risk-neutral probability measure $\mathbb{Q}$. For $D = 2$, the dynamics for our multi-dimensional Heston model are given by

$$\begin{aligned} dV_1(t) &= \kappa_1(\theta_1 - V_1(t))dt + \varepsilon_1\sqrt{V_1(t)}dW_{V_1}(t), \\ dV_2(t) &= \kappa_2(\theta_2 - V_2(t))dt + \varepsilon_2\sqrt{V_2(t)}dW_{V_2}(t), \\ dS_1(t) &= S_1(t)r(t)dt + S_1(t)\sqrt{V_1(t)}dW_{S_1}(t), \\ dS_2(t) &= S_2(t)r(t)dt + S_2(t)\sqrt{V_2(t)}dW_{S_2}(t), \end{aligned}$$

where $W_{V_i}(t), W_{S_i}(t)$ are Brownian motions under $\mathbb{Q}$ for $i = 1, 2$. For now, we assume the following correlation structure:

$$\Sigma = \begin{pmatrix} 1 & \rho_{V_1V_2} & \rho_1 & 0 \\ \rho_{V_1V_2} & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & \rho_{12} \\ 0 & \rho_2 & \rho_{12} & 1 \end{pmatrix},$$

Compared to the previous sections, we drop the assumption of zero variance-variance correlation. Note that each of these stocks, priced in terms of the money market, is a martingale with respect to $\mathbb{Q}$ and filtration $\mathcal{F}(t)$ up to time t , since

$$\mathbb{E} \left[\frac{S_i(T)}{B(T)} \middle| \mathcal{F}(t) \right] = \frac{S_i(t)}{B(t)} + \mathbb{E} \left[\int_t^T d \left(\frac{S_i(s)}{B(s)} \right) \middle| \mathcal{F}(t) \right] = \frac{S_i(t)}{B(t)} \quad \forall t \leq T, i = 1, 2.$$

For the most right equality, we used Itô's Lemma on the function $f(t, x) = xB^{-1}(t)$ to find

$$d \left(\frac{S_i(s)}{B(s)} \right) = \left(-\frac{S_i(s)r(s)}{B(s)} + \frac{S_i(s)r(s)}{B(s)} + 0 \right) dt + \frac{S_i(s)\sqrt{V_i(s)}}{B(s)} dW_{S_i}(s) = \frac{S_i(s)\sqrt{V_i(s)}}{B(s)} dW_{S_i}(s),$$

which implies that, given $\mathcal{F}(t)$,

$$\int_t^T d \left(\frac{S_i(s)}{B(s)} \right) = \int_t^T \frac{\sqrt{V_i(s)}}{B(s)} dW_{S_i}(s)$$

is an Itô integral, which has expectation zero. Therefore, $S_i(t)$ is indeed a $\mathbb{Q}$ -martingale for $i = 1, 2$.

5.4.2 Strategy outline in case of Black-Scholes dynamics

Our strategy to price one specific basket option semi-analytically is based on an existing strategy, as described in [Björk, 2009]. In this strategy, both stocks are assumed to have Black-Scholes dynamics. We will explain this strategy briefly to give an idea of our own strategy.

Björk assumes that $S_1(t)$ and $S_2(t)$ follow Black-Scholes dynamics:

$$\begin{aligned} dS_1(t) &= S_1(t)rdt + S_1(t)\sigma_1dW_{S_1}(t), \\ dS_2(t) &= S_2(t)rdt + S_2(t)\sigma_2dW_{S_2}(t), \end{aligned}$$

¹A numéraire is the unit of account in which other assets are denominated.

with σ_1, σ_2 positive constants. Now consider an *exchange option* with a payoff given by $\max[S_2(T) - KS_1(T), 0]$. One can price this exchange option by use of a Monte Carlo method. In order to find a reference price, Björk changes the numéraire from the money market account $B(t)$ to $S_1(t)$. He defines the Radon-Nikodým derivative

$$\frac{d\mathbb{S}}{d\mathbb{Q}} = \frac{B(t)S_1(T)}{B(T)S_1(t)}. \quad (5.3)$$

Here, $\mathbb{S}$ is the induced measure, and Björk shows that the exchange option value at time t can be written as

$$\frac{S_1(t)}{B(t)} \mathbb{E}^{\mathbb{S}} [\max(S_2(T) - K, 0) | \mathcal{F}(t)],$$

with $S_q(t) := S_2(t)/S_1(t)$. This means that the problem reduces to the validation of a European call option under $\mathbb{S}$. To price this option, one needs the dynamics of $S_q(t)$ under $\mathbb{S}$. Björk finds these dynamics via a Girsanov transformation, and shows that $S_q(t)$ is an $\mathbb{S}$ -martingale. Since $S_q(T)$ is log-normally distributed given $S(t)$, one can use the Black-Scholes formula to price the call option. In this way, we find the analytic price of the exchange option.

Our strategy is analogous to the above strategy. Of course, the Black-Scholes formula is not useful for us since we assume Heston dynamics. However, in one dimension, a semi-analytic call option price can be derived when the underlying follows Heston dynamics. The analogy between the Black-Scholes strategy and our Heston strategy will not be straightforward. Therefore, we will be strict in each derivation step in the following subsections.

5.4.3 Change of numéraire under Heston

Let us return to our assumption that $S_1(t)$ and $S_2(t)$ follow Heston dynamics and change the numéraire similarly to the method in [Björk, 2009]. We will change the numéraire from $B(t)$ to $S_1(t)$. We define the probability measure $\mathbb{S}$ by the Radon-Nikodým derivative:

$$\frac{d\mathbb{S}}{d\mathbb{Q}} = \frac{B(t)S_1(T)}{B(T)S_1(t)}. \quad (5.4)$$

Since $B(t)$ is an exponential function and $S_1(0) > 0$ by assumption, $B(t), S_1(t) > 0 \ \forall t \geq 0$ almost surely. Therefore, the fraction in (5.4) is almost surely strictly positive for all t . Furthermore, since $S_1(t)/B(t)$ is a $\mathbb{Q}$ -martingale, we have

$$\mathbb{E} \left[\frac{d\mathbb{S}}{d\mathbb{Q}} \middle| \mathcal{F}(t) \right] = \frac{B(t)}{S_1(t)} \mathbb{E} \left[\frac{S_1(T)}{B(T)} \middle| \mathcal{F}(t) \right] = \frac{B(t)}{S_1(t)} \frac{S_1(t)}{B(t)} = 1.$$

This shows that the probability measure $\mathbb{S}$ is well-defined (see [Shreve, 2004]). Therefore, for every non-negative random variable X , we can write

$$\mathbb{E}^{\mathbb{S}}[X] = \mathbb{E} \left[X \frac{d\mathbb{S}}{d\mathbb{Q}} \right]. \quad (5.5)$$

This yields, for all $0 \leq t \leq T$,

$$\mathbb{E}^{\mathbb{S}} \left[\frac{S_2(T)}{S_1(T)} \middle| \mathcal{F}(t) \right] = \mathbb{E} \left[\frac{S_2(T)}{S_1(T)} \frac{B(t)S_1(T)}{B(T)S_1(t)} \middle| \mathcal{F}(t) \right] = \frac{B(t)}{S_1(t)} \mathbb{E} \left[\frac{S_2(T)}{B(T)} \middle| \mathcal{F}(t) \right] = \frac{B(t)}{S_1(t)} \frac{S_2(t)}{B(t)} = \frac{S_2(t)}{S_1(t)}.$$

In the second last equality, we used the fact that $S_2(t)/B(t)$ is a martingale under $\mathbb{Q}$. We conclude that $S_2(T)/S_1(T)$ is a martingale under measure $\mathbb{S}$.

Now let us consider an exchange option, which gives the holder the right to exchange one S_2 -share for K S_1 -shares at time T . We recall that the payoff is $\max[S_2(T) - KS_1(T), 0]$. The option value can be written as

$$\mathbb{E} [\max(S_2(T) - KS_1(T), 0) / B(T) | \mathcal{F}(t)]. \quad (5.6)$$

In order to change the measure, we first note that since $d\mathbb{S}/d\mathbb{Q}$ is almost surely strictly positive, we can transform (5.5) to

$$\mathbb{E}[X] = \mathbb{E}^{\mathbb{S}}\left[X \frac{d\mathbb{Q}}{d\mathbb{S}}\right].$$

Therefore, we can write (5.6) as

$$\mathbb{E}^{\mathbb{S}}\left[\max(S_2(T) - KS_1(T), 0) \frac{B(T)S_1(t)}{B(t)B(T)S_1(T)} \middle| \mathcal{F}(t)\right] = \frac{S_1(t)}{B(t)} \mathbb{E}^{\mathbb{S}}[\max(S_q(T) - K, 0) | \mathcal{F}(t)],$$

with $S_q(t) := S_2(t)/S_1(t)$ (the index q refers to the quotient of the two stocks). To conclude, the problem reduces to the validation of a European call option on the $\mathbb{S}$ -martingale $S_q(t)$ with strike price K .

5.4.4 Dynamics of $S_q(t)$ under $\mathbb{S}$

In order to price this European call option, we first have to find the $\mathbb{S}$ -dynamics of $S_q(t)$. We will use Itô's Lemma in two dimensions:

Result 4. Itô's Lemma, two dimensions

Let $f(t, x, y)$ be a function whose partial derivatives $f_t, f_x, f_y, f_{xx}, f_{yy}, f_{xy}$ are defined and are continuous. Let $X(t)$ and $Y(t)$ be Itô processes. Then

$$\begin{aligned} df(t, X(t), Y(t)) &= \frac{\partial f}{\partial t}(t, X(t), Y(t))dt + \frac{\partial f}{\partial x}(t, X(t), Y(t))dX(t) + \frac{\partial f}{\partial y}(t, X(t), Y(t))dY(t) \\ &\quad + \frac{1}{2} \begin{pmatrix} dX(t) & dY(t) \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(t, X(t), Y(t)) & \frac{\partial^2 f}{\partial x \partial y}(t, X(t), Y(t)) \\ \frac{\partial^2 f}{\partial x \partial y}(t, X(t), Y(t)) & \frac{\partial^2 f}{\partial y^2}(t, X(t), Y(t)) \end{pmatrix} \begin{pmatrix} dX(t) \\ dY(t) \end{pmatrix}. \end{aligned}$$

The application of Itô's Lemma in two dimensions on the function $f(t, x, y) = y/x$ yields

$$\begin{aligned} dS_q(t) &= df(t, S_1(t), S_2(t)) \\ &= 0 - \frac{S_2(t)}{S_1^2(t)} \left(S_1(t)r(t)dt + S_1(t)\sqrt{V_1(t)}dW_{S_1}(t) \right) + \frac{1}{S_1(t)} \left(S_2(t)r(t)dt + S_2(t)\sqrt{V_2(t)}dW_{S_2}(t) \right) \\ &\quad + \frac{1}{2} \left(\frac{2S_2(t)}{S_1^3(t)} (dS_1(t))^2 - \frac{1}{S_1^2(t)} dS_1(t)dS_2(t) - \frac{1}{S_1^2(t)} dS_1(t)dS_2(t) + 0 \right) \\ &= S_q(t) \left(\sqrt{V_2(t)}dW_{S_2}(t) - \sqrt{V_1(t)}dW_{S_1}(t) \right) + \frac{S_2(t)}{S_1^3(t)} (dS_1(t))^2 - \frac{1}{S_1^2(t)} dS_1(t)dS_2(t). \end{aligned} \quad (5.7)$$

If we write out $dS_i(t)dS_j(t)$ for $i, j = 1, 2$, all terms with dt^2 and $dW_{S_i}(t)dt$ vanish since we neglect higher order terms. This yields

$$dS_i(t)dS_j(t) = S_i(t)S_j(t)\rho_{ij}\sqrt{V_i(t)}\sqrt{V_j(t)}dt = S_i(t)S_j(t)\rho_{ij}\sqrt{V_i(t)V_j(t)}dt, \quad \text{for } i, j = 1, 2.$$

In the last equality, we used the fact that both variance paths are non-negative for all t . Substituting this formula in (5.7), yields

$$dS_q(t) = S_q(t) \left(V_1(t) - \rho_{12}\sqrt{V_1(t)V_2(t)} \right) dt + S_q(t) \left(\sqrt{V_2(t)}dW_{S_2}(t) - \sqrt{V_1(t)}dW_{S_1}(t) \right). \quad (5.8)$$

Now we have arrived at the one-dimensional $\mathbb{Q}$ -dynamics of $S_q(t)$. However, we desire the $\mathbb{S}$ -dynamics of $S_q(t)$. Therefore, we will apply the multi-dimensional version of Girsanov's Theorem:

Result 5. Girsanov's Theorem, multiple dimensions

Let $\mathbf{W}(t), 0 \leq t \leq T$, be a d -dimensional Brownian motion vector with independent components on a probability space

$(\Omega, \mathcal{F}, \mathbb{P})$. - Let $\mathcal{F}(t), 0 \leq t \leq T$, be the filtration generated by this Brownian motion vector. Let $\mathbf{\Theta}(t)$ be a d -dimensional adapted process. Define

$$\begin{aligned} Z(t) &= e^{-\int_0^t \mathbf{\Theta}^\top(s) d\mathbf{W}(s) - \frac{1}{2} \int_0^t \|\mathbf{\Theta}(s)\|^2 ds}, \\ \widetilde{\mathbf{W}}(t) &= \mathbf{W}(t) + \int_0^t \mathbf{\Theta}(s) ds, \end{aligned} \quad (5.9)$$

where $\|\mathbf{x}\|$ denotes the Euclidean norm of $\mathbf{x}$. Assume that

$$\mathbb{E} \int_0^T \|\mathbf{\Theta}(s)\|^2 Z^2(s) ds < \infty. \quad (5.10)$$

Then $\mathbb{E}[Z(T)] = 1$, and under the probability measure $\widetilde{\mathbb{P}}$, given by

$$\widetilde{\mathbb{P}}(A) = \int_A Z(\omega) d\mathbb{P}(\omega) \quad \forall A \in \mathcal{F}(t),$$

the process $\widetilde{\mathbf{W}}(t)$ is a d -dimensional Brownian motion.

Note that the theorem uses *independent* Brownian motions, whereas our model has a correlation structure given by the 4×4 correlation matrix Σ . Therefore, we will transform our model: suppose that L is a lower triangular matrix so that $\Sigma = LL^\top$. Then we can write

$$\begin{pmatrix} dW_{V_1}(t) \\ dW_{V_2}(t) \\ dW_{S_1}(t) \\ dW_{S_2}(t) \end{pmatrix} \stackrel{d}{=} L \begin{pmatrix} dW_1(t) \\ dW_2(t) \\ dW_3(t) \\ dW_4(t) \end{pmatrix}, \quad (5.11)$$

in which all components of the right-hand side vector are independent Brownian motion increments. Substitution in (5.8) yields the $\mathbb{Q}$ -dynamics of $S_q(t)$:

$$\begin{aligned} \frac{dS_q(t)}{S_q(t)} &= \left(V_1(t) - \rho_{12} \sqrt{V_1(t)V_2(t)} \right) dt + \left(\sqrt{V_2(t)} L_{41} - \sqrt{V_1(t)} L_{31} \right) dW_1(t) \\ &\quad + \left(\sqrt{V_2(t)} L_{42} - \sqrt{V_1(t)} L_{32} \right) dW_2(t) + \left(\sqrt{V_2(t)} L_{43} - \sqrt{V_1(t)} L_{33} \right) dW_3(t) \\ &\quad + \sqrt{V_2(t)} L_{44} dW_4(t). \end{aligned} \quad (5.12)$$

We apply the theorem for $\mathbb{P} = \mathbb{Q}$, $\widetilde{\mathbb{P}} = \mathbb{S}$ and $Z(T)$ equal to the proposed Radon-Nikodým derivative in (5.3). We rewrite the integrated form of (2.3), to obtain

$$S_1(t) = S_1(0) e^{\int_0^t r(s) ds - \frac{1}{2} \int_0^t V_1(s) ds + \int_0^t \sqrt{V_1(s)} dW_{S_1}(s)},$$

which yields

$$Z(T) = \left. \frac{d\mathbb{S}}{d\mathbb{Q}} \right|_{t=0} = \frac{B(0)S_1(T)}{B(T)S_1(0)} = e^{-\frac{1}{2} \int_0^T V_1(s) ds + \int_0^T \sqrt{V_1(s)} dW_{S_1}(s)}.$$

Comparing this expression with (5.9), we conclude that we have to define $\mathbf{\Theta}(t)$ as follows:

$$\mathbf{\Theta}(t) = \begin{pmatrix} 0 \\ 0 \\ -\sqrt{V_1(t)} \\ 0 \end{pmatrix}.$$

Note that this is the only possible choice for $\mathbf{\Theta}(t)$. To validate condition (5.10), we note that

$$\mathbb{E} \int_0^T \|\mathbf{\Theta}(s)\|^2 Z^2(s) ds = \mathbb{E} \int_0^T V_1(s) \frac{B^2(0)S_1^2(s)}{B^2(s)S_1^2(0)} ds = \frac{B^2(0)}{S_1^2(0)} \int_0^T \mathbb{E}[V_1(s)S_1^2(s)] B^{-2}(s) ds, \quad (5.13)$$

where we used Fubini's Theorem in the second equality. The moment $\mathbb{E}[V_1(s)S_1^2(s)]$ can be computed by using results in Appendix A of [Andersen, 2007]. There, the analytic expression of the joint characteristic function of $V_1(T)$ and $X_1(T)$, $\phi(u, v) = \mathbb{E}[e^{iuV_1(T)+iv(X_1(T)-X_1(0))}]$, is derived (which we will not repeat here). Then, using the fact that

$$\frac{\partial \phi}{\partial u}(0, -2i) = \mathbb{E}[iV_1(T)e^{2(X_1(T)-X_1(0))}] = ie^{-2X(0)}\mathbb{E}[V_1(T)S_1^2(T)],$$

we can derive the left-hand side (which we will not write out). In this way, we derived that the right-hand side of (5.13) is finite. We conclude that it is appropriate to apply Girsanov's Theorem. Now recall that we restrict that $S_q(t)$ is an $\mathbb{S}$ -martingale. Loosely speaking, all dt -terms should vanish by applying Girsanov's Theorem. Therefore, by substituting $dW_i(t) = dW_i^{\mathbb{S}}(t) - \Theta_i(t)dt$ for $i = 1, 2, 3, 4$ in (5.12), it is straightforward that

$$V_1(t) - \rho_{12}\sqrt{V_1(t)V_2(t)} = -\left(\sqrt{V_2(t)}L_{43} - \sqrt{V_1(t)}L_{33}\right)\sqrt{V_1(t)}, \text{ for } t \geq 0, \quad (5.14)$$

must hold. This implies that one must have $L_{43} = \rho_{12}$ and $L_{33} = 1$. However, a Cholesky decomposition of Σ (see Example 1 of Appendix A), shows that $L_{33} = 1$ if and only if $\rho_1 = 0$. If this is the case, then $L_{43} = \rho_{12}$ follows immediately. We conclude that the assumption $\rho_1 = 0$ is necessary. Then $S_q(t)$ is a $\mathbb{S}$ -martingale, with $\mathbb{S}$ -dynamics

$$\begin{aligned} \frac{dS_q(t)}{S_q(t)} &= \left(\sqrt{V_2(t)}L_{41} - \sqrt{V_1(t)}L_{31}\right)dW_1^{\mathbb{S}}(t) + \left(\sqrt{V_2(t)}L_{42} - \sqrt{V_1(t)}L_{32}\right)dW_2^{\mathbb{S}}(t) \\ &\quad + \left(\sqrt{V_2(t)}L_{43} - \sqrt{V_1(t)}L_{33}\right)dW_3^{\mathbb{S}}(t) - \sqrt{V_1(t)}L_{44}dW_4^{\mathbb{S}}(t) \\ &= \sqrt{V_2(t)}\left(L_{41}dW_1^{\mathbb{S}}(t) + L_{42}dW_2^{\mathbb{S}}(t) + L_{43}dW_3^{\mathbb{S}}(t) + L_{44}dW_4^{\mathbb{S}}(t)\right) \\ &\quad - \sqrt{V_1(t)}\left(L_{31}dW_1^{\mathbb{S}}(t) + L_{32}dW_2^{\mathbb{S}}(t) + L_{33}dW_3^{\mathbb{S}}(t)\right). \end{aligned} \quad (5.15)$$

The correlation structure between Brownian motions remains unchanged after a change of measure:

$$dW_i(t)dW_j(t) = (dW_i^{\mathbb{S}}(t) - \Theta_i(t)dt)(dW_j^{\mathbb{S}}(t) - \Theta_j(t)dt) = dW_i^{\mathbb{S}}(t)dW_j^{\mathbb{S}}(t).$$

Therefore, the correlation matrix of the $\mathbb{S}$ -Brownian motions is Σ and we can use the same Cholesky decomposition $\Sigma = LL^{\top}$. We substitute (5.11) in (5.15) to end up with a more simple expression of the $\mathbb{S}$ -dynamics of $S_q(t)$:

$$\frac{dS_q(t)}{S_q(t)} = \sqrt{V_2(t)}dW_{S_2}^{\mathbb{S}}(t) - \sqrt{V_1(t)}dW_{S_1}^{\mathbb{S}}(t). \quad (5.16)$$

The derivation of the $\mathbb{S}$ -dynamics of $V_1(t)$ and $V_2(t)$ is straightforward. Note that

$$dW_{V_i}(t) = dW_i(t) = dW_i^{\mathbb{S}}(t) - \Theta_i(t)dt = dW_i^{\mathbb{S}}(t) = dW_{V_i}^{\mathbb{S}}(t), \quad i = 1, 2.$$

Here, the first and last equality are due to the independence of the variance processes under respective measures $\mathbb{Q}$ and $\mathbb{S}$, the second equality represents the application of Girsanov, and the third equality holds by choice of $\Theta(t)$. We conclude that the dynamics of the variance processes under $\mathbb{S}$ are similar to those under $\mathbb{Q}$:

$$\begin{aligned} dV_1(t) &= \kappa_1(\theta_1 - V_1(t))dt + \varepsilon_1\sqrt{V_1(t)}dW_{V_1}^{\mathbb{S}}(t), \\ dV_2(t) &= \kappa_2(\theta_2 - V_2(t))dt + \varepsilon_2\sqrt{V_2(t)}dW_{V_2}^{\mathbb{S}}(t). \end{aligned}$$

5.4.5 Conditions for model affinity

Now that we have found the $\mathbb{S}$ -dynamics of $S_q(t)$, we have to price a European call option with $S_q(t)$ as underlying. As Björk assumes Black-Scholes dynamics, he is able to use the Black-Scholes formula to obtain the analytic price of this option. In our model, since we assume Heston dynamics, we can instead compute a semi-analytic price by use of a Fourier method.

In order to find the Fourier price of the European call option on $S_q(t)$, we must find the characteristic function of $\ln S_q(t)$. For affine models, this characteristic function can be derived by following a specific procedure, as described in [Duffie et al., 2000]. We do not yet know whether our model is affine. Let us first introduce the definition of this property:

Definition. Suppose $\mathbf{X}(t) \in \mathbb{R}^N$ is the vector of state variables of an N -dimensional SDE, and suppose we can write its dynamics as

$$d\mathbf{X} = \mu(t, \mathbf{X}(t))dt + \sigma(\mathbf{X}(t))d\mathbf{W}(t),$$

with $\mathbf{W}(t) \in \mathbb{R}^{2D}$ a standard Brownian motion vector with independent components, $\mu(t, \mathbf{X}(t)) \in \mathbb{R}^N$ the drift component and $\sigma(\mathbf{X}(t)) \in \mathbb{R}^{N \times 2D}$ the diffusion component. Then the model is affine if

$$\mu(t, \mathbf{x}) = K_0(t) + K_1 \mathbf{x}, \quad K_0(t) \in \mathbb{R}^N, K_1 \in \mathbb{R}^{N \times N}, \quad (5.17)$$

$$\left[\sigma(\mathbf{x}) \sigma(\mathbf{x})^\top \right]_{ij} = [H_0]_{ij} + [H_1]_{ij} \mathbf{x} \quad \forall i, j = 1, 2, \dots, N, \quad H_0 \in \mathbb{R}^{N \times N}, H_1 \in \mathbb{R}^{N \times N \times N}. \quad (5.18)$$

In these expressions, $K_0(t)$ is a function of time, K_1 , H_0 and H_1 are constants, and $[H_1]_{ij}$ is a $1 \times N$ -vector component of H_1 .

Let us consider $f(t, S_q(t)) = \ln S_q(t)$. Applying Itô's lemma on this function, using (5.16), yields

$$dX_q(t) := d \ln S_q(t) = -\frac{1}{2} (V_2(t) + V_1(t) - 2\rho_{12} \sqrt{V_1(t)V_2(t)}) dt + \sqrt{V_2(t)} dW_{S_2}^S(t) - \sqrt{V_1(t)} dW_{S_1}^S(t).$$

Furthermore, let us define

$$\mathbf{X}(t) = \begin{pmatrix} V_1(t) \\ V_2(t) \\ X_q(t) \end{pmatrix}, \mathbf{W}(t) = \begin{pmatrix} W_1(t) \\ W_2(t) \\ W_3(t) \\ W_4(t) \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & \rho_{V_1 V_2} & \rho_1 & 0 \\ \rho_{V_1 V_2} & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & \rho_{12} \\ 0 & \rho_2 & \rho_{12} & 1 \end{pmatrix},$$

with $\mathbf{W}(t)$ a vector of independent $\mathbb{S}$ -Brownian motions. Then we know that we can write

$$\begin{pmatrix} W_{V_1}^S(t) \\ W_{V_2}^S(t) \\ W_{S_1}^S(t) \\ W_{S_2}^S(t) \end{pmatrix} \stackrel{d}{=} L \mathbf{W}(t),$$

with L the lower triangular matrix obtained by a Cholesky decomposition of the correlation matrix $\Sigma = LL^\top$ of the Brownian motion vector on the left-hand side. Therefore, we can write the model in the following vector form:

$$d\mathbf{X}(t) = \mu(t, \mathbf{X}(t))dt + S(\mathbf{X}(t))Ld\mathbf{W}(t),$$

with

$$\mu(t, \mathbf{X}(t)) = \begin{pmatrix} \kappa_1(\theta_1 - V_1(t)) \\ \kappa_2(\theta_2 - V_2(t)) \\ -\frac{1}{2} (V_2(t) + V_1(t) - 2\rho_{12} \sqrt{V_1(t)V_2(t)}) \end{pmatrix},$$

$$S(\mathbf{X}(t)) = \begin{pmatrix} \varepsilon_1 \sqrt{V_1(t)} & 0 & 0 & 0 \\ 0 & \varepsilon_2 \sqrt{V_2(t)} & 0 & 0 \\ 0 & 0 & -\sqrt{V_1(t)} & \sqrt{V_2(t)} \end{pmatrix}.$$

Let us investigate for which parameter sets, $\{\kappa_1, \theta_1, \varepsilon_1, V_1(0), \kappa_2, \theta_2, \varepsilon_2, V_2(0), \Sigma\}$, this model is affine. The drift component, $\mu(t, \mathbf{X}(t))$, is not linear in $\mathbf{X}(t)$ in general. Furthermore, for the diffusion component times its transpose, we find

$$\begin{aligned} \sigma(\mathbf{X}(t)) \sigma^\top(\mathbf{X}(t)) &= S(\mathbf{X}(t))L(S(\mathbf{X}(t))L)^\top = S(\mathbf{X}(t))LL^\top S(\mathbf{X}(t))^\top = S(\mathbf{X}(t))\Sigma S(\mathbf{X}(t))^\top \\ &= \begin{pmatrix} \varepsilon_1^2 V_1(t) & \rho_{V_1 V_2} \varepsilon_1 \varepsilon_2 \sqrt{V_1(t)V_2(t)} & -\varepsilon_1 \rho_1 V_1(t) \\ \rho_{V_1 V_2} \varepsilon_1 \varepsilon_2 \sqrt{V_1(t)V_2(t)} & \varepsilon_2^2 V_2(t) & \varepsilon_2 \rho_2 V_2(t) \\ -\varepsilon_1 \rho_1 V_1(t) & \varepsilon_2 \rho_2 V_2(t) & V_1(t) + V_2(t) - 2\rho_{12} \sqrt{V_2(t)V_1(t)} \end{pmatrix}. \end{aligned}$$

Before we will write the above expression in terms of H_0 and H_1 , note that only three elements of the right-hand side matrix are non-affine in $\mathbf{X}(t)$ for all Θ . In those three elements as well as in the drift component, the square-root of $V_1(t)V_2(t)$ prevents the model from being affine. We can eliminate this problem in two ways:

1. Firstly, we can set $\rho_{12} = \rho_{V_1 V_2} = 0$, but this would imply that the problem would tantamount to two independent one-dimensional Heston models. However, since the uncertainty of the performance of our new model lies in the very fact that it is multi-dimensional, this solution would obviously be unacceptable.
2. Alternatively, we can choose the two variance dynamics proportional to each other:

$$b^2 V_2(t) = V_1(t),$$

for a constant $b \in \mathbb{R}$. We use the square of b for notational convenience. Note that this assumption allows $V_1(t) = 0$ for all t . In this case, the dynamics of $S_1(t)$ are deterministic. $V_2(t) = 0$ is not allowed. Nevertheless, by a symmetry argument, we have not lost generality. Since both variance dynamics are non-negative, the equality $b\sqrt{V_1(t)V_2(t)} = \pm V_1(t)$ holds and we can thus obtain an affine diffusion component.

Hence we conclude that the considered model is affine if we assume $b^2 V_2(t) = V_1(t)$.

5.4.6 Derivation of the characteristic function

Let us make this assumption and derive the characteristic function of the model. The assumption reduces our model to a two-dimensional model with three Brownian motions:

$$\begin{aligned} dV_2(t) &= \kappa_2(\theta_2 - V_2(t))dt + \varepsilon_2\sqrt{V_2(t)}dW_{V_2}^{\mathbb{S}}(t), \\ dX_q(t) &= -\frac{1}{2}(b^2 - 2b\rho_{12} + 1)V_2(t)dt - b\sqrt{V_2(t)}dW_{S_1}^{\mathbb{S}}(t) + \sqrt{V_2(t)}dW_{S_2}^{\mathbb{S}}(t). \end{aligned}$$

Note that for $b = 0$, the model reduces to a one-dimensional Heston model with zero mean rate of return. We can write this affine model in vector form

$$d\mathbf{X}(t) = \mu(t, \mathbf{X}(t))dt + S(\mathbf{X}(t))Ld\mathbf{W}(t),$$

by setting

$$\begin{aligned} \mathbf{X}(t) &= \begin{pmatrix} V_2(t) \\ X_q(t) \end{pmatrix}, \mu(t, \mathbf{X}(t)) = K_0 + K_1\mathbf{X}(t), K_0 = \begin{pmatrix} \kappa_2\theta_2 \\ 0 \end{pmatrix}, K_1 = \begin{pmatrix} -\kappa_2 & 0 \\ -(b^2 - 2b\rho_{12} + 1)/2 & 0 \end{pmatrix}, \\ S(\mathbf{X}(t)) &= \begin{pmatrix} \varepsilon_2\sqrt{V_2(t)} & 0 & 0 \\ 0 & -b\sqrt{V_2(t)} & \sqrt{V_2(t)} \end{pmatrix}, \mathbf{W}(t) = \begin{pmatrix} W_1(t) \\ W_2(t) \\ W_3(t) \end{pmatrix}. \end{aligned}$$

The elements of $\mathbf{W}(t)$ are independent $\mathbb{S}$ -Brownian motions, and L is the lower triangular matrix of a Cholesky decomposition of the correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_{12} \\ \rho_2 & \rho_{12} & 1 \end{pmatrix}$$

of the Brownian motion vector

$$\begin{pmatrix} W_{V_2}^{\mathbb{S}}(t) \\ W_{S_1}^{\mathbb{S}}(t) \\ W_{S_2}^{\mathbb{S}}(t) \end{pmatrix}. \quad (5.19)$$

The diffusion component becomes

$$\sigma(\mathbf{X}(t))\sigma^\top(\mathbf{X}(t)) = S(\mathbf{X}(t))\Sigma S(\mathbf{X}(t))^\top = \begin{pmatrix} \varepsilon_2^2 V_2(t) & \varepsilon_2(\rho_2 - b\rho_1)V_2(t) \\ \varepsilon_2(\rho_2 - b\rho_1)V_2(t) & (b^2 - 2b\rho_{12} + 1)V_2(t) \end{pmatrix},$$

so that the diffusion component can be written in the form as given in (5.18), by taking $H_0 = 0$ and H_1 a $2 \times 2 \times 2$ -matrix which we will express in parts along its third dimension:

$$(H_1)_1 = \begin{pmatrix} \varepsilon_2^2 & \varepsilon_2(\rho_2 - b\rho_1) \\ \varepsilon_2(\rho_2 - b\rho_1) & b^2 - 2b\rho_{12} + 1 \end{pmatrix}, (H_1)_2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now we will use an important result in [Duffie et al., 2000]. Since the model is affine, the characteristic function of the vector-valued SDE is of the form

$$\phi(u, \mathbf{x}, t, T) = e^{\alpha(t) + \mathbf{x}\beta(t)},$$

with $u \in \mathbb{C}$ and $\mathbf{x} = \begin{pmatrix} v & x_q \end{pmatrix}$. Furthermore, $\alpha : [0, \infty) \rightarrow \mathbb{R}$ and $\beta : [0, \infty) \rightarrow \mathbb{R}^2$ satisfy the complex-valued ordinary differential equations²

$$\begin{aligned} \frac{d\beta}{dt}(t) &= -K_1^\top \beta(t) - \frac{1}{2} \beta(t)^\top H_1 \beta(t), \\ \frac{d\alpha}{dt}(t) &= -K_0^\top \beta(t) - \frac{1}{2} \beta(t)^\top H_0 \beta(t), \end{aligned}$$

with boundary conditions $\alpha(T) = 0$ and $\beta(T) = \begin{bmatrix} 0 & iu \end{bmatrix}^\top \forall u \in \mathbb{C}$, where i is the imaginary unit. In our model, $\alpha(t)$ and $\beta(t)$ must satisfy:

$$\begin{aligned} \frac{d\beta_1(t)}{dt} &= \frac{1}{2}((b^2 - 2b\rho_{12} + 1)\beta_2(t) \\ &\quad - (b^2 - 2b\rho_{12} + 1)\beta_2^2(t)) + (\kappa_2 - \varepsilon_2(\rho_2 - b\rho_1)\beta_2(t))\beta_1(t) - \frac{\varepsilon_2^2}{2}\beta_1^2(t), \\ \frac{d\beta_2(t)}{dt} &= 0, \\ \frac{d\alpha(t)}{dt} &= -\kappa_2\theta_2\beta_1(t). \end{aligned}$$

To obtain a more familiar ODE, we perform a change of variables $\tau := T - t$, the time up to maturity, and we solve for $\beta(\tau)$ and $\alpha(\tau)$. This will only change the sign of the right-hand side of the three ODEs and the boundary conditions:

$$\begin{aligned} \frac{d\beta_1(\tau)}{d\tau} &= -\frac{1}{2}((b^2 - 2b\rho_{12} + 1)\beta_2(\tau) - (b^2 - 2b\rho_{12} + 1)\beta_2^2(\tau)) \\ &\quad - (\kappa_2 - \varepsilon_2(\rho_2 - b\rho_1)\beta_2(\tau))\beta_1(\tau) + \frac{\varepsilon_2^2}{2}\beta_1^2(\tau), \end{aligned} \quad (5.20)$$

$$\frac{d\beta_2(\tau)}{d\tau} = 0, \quad (5.21)$$

$$\frac{d\alpha(\tau)}{d\tau} = \kappa_2\theta_2\beta_1(\tau). \quad (5.22)$$

with $\alpha(0) = 0$, $\beta_1(0) = 0$ and $\beta_2(0) = iu \forall u \in \mathbb{C}$. This last boundary condition combined with (5.21) immediately implies $\beta_2(\tau) = iu$. Substituting this in the first of the three ODEs yields

$$\frac{d\beta_1(\tau)}{d\tau}(\tau) = -\frac{1}{2}((b^2 - 2b\rho_{12} + 1)iu + (b^2 - 2b\rho_{12} + 1)u^2) - (\kappa_2 - \varepsilon_2(\rho_2 - b\rho_1)iu)\beta_1(\tau) + \frac{\varepsilon_2^2}{2}\beta_1^2(\tau),$$

which we recognize as a Ricatti equation. The general solution of a Ricatti equation, given by

$$\frac{dg(\tau)}{d\tau} = c_0 + c_1g(\tau) + c_2g^2(\tau),$$

with boundary condition $g(0) = g_0$, is

$$g(\tau) = g_0 + \frac{(-c_1 - d - 2c_2g_0)(1 - e^{-d\tau})}{2c_2(1 - Ge^{-d\tau})}, \text{ where } d = \sqrt{c_1^2 - 4c_0c_2} \text{ and } G = \frac{-c_1 - d - 2g_0c_2}{-c_1 + d - 2g_0c_2}.$$

Now, we can find $\beta_1(\tau)$, by setting $c_0 = -\frac{1}{2}((b^2 - 2b\rho_{12} + 1)iu + (b^2 - 2b\rho_{12} + 1)u^2)$, $c_1 = -(\kappa_2 - \varepsilon_2(\rho_2 - b\rho_1)iu)$, $c_2 = \varepsilon_2^2/2$ and $g_0 = 0$. The solution of the last ODE, (5.22), is then given by

$$\alpha(\tau) = \frac{\kappa_2\theta_2}{2c_2} \left((-c_1 - d)\tau - 2 \log \frac{1 - Ge^{-d\tau}}{1 - G} \right).$$

We have fully specified the characteristic function, and we can thus find a reference value for the exchange option by use of a Fourier method.

²Here, $c^\top H_1 c$ denotes the the vector in $\mathbb{C}^n$ with k -th element $\sum_{i,j} c_i(H_1)_{ijk}c_j$.

5.4.7 Results

We implement the above in MatLab, choosing the parameters so that the Feller condition was violated for both stocks.

	S_1	S_2
$S_i(0)$	100	100
$r(t)$	0	0
κ_i	0.5	0.5
$\theta_i, V_i(0)$	0.16	0.04
ε_i	2	1

Table 5.3: Parameter set of stocks S_1 and S_2

It is shown in Table 5.3 that stock S_2 has the same parameter value as in *Feller Violated* in Section 5.1. Furthermore, the parameters of stock S_1 are chosen so that the model satisfies one of our derived conditions to be able to find a reference value for the exchange option: $V_1(t) = b^2 V_2(t)$, $b = 2$. We choose $T = 1$. The other conditions ($\rho_1 = 0, \rho_{V_1 V_2} = 1$) are satisfied by choosing a correlation structure given by

$$\Sigma = \begin{pmatrix} 1 & \rho_{V_1 V_2} & \rho_1 & \rho_{V_1 S_2} \\ \rho_{V_1 V_2} & 1 & \rho_{V_2 S_1} & \rho_2 \\ \rho_1 & \rho_{V_2 S_1} & 1 & \rho_{12} \\ \rho_{V_1 S_2} & \rho_2 & \rho_{12} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & -0.6 \\ 1 & 1 & 0 & -0.6 \\ 0 & 0 & 1 & 0.7 \\ -0.6 & -0.6 & 0.7 & 1 \end{pmatrix}.$$

At first, we choose a large number of Monte Carlo paths and steps³, to show that the MQE scheme and the Fourier method agree on the exchange option price. One can see in Figure 5.5 that this is indeed the case.

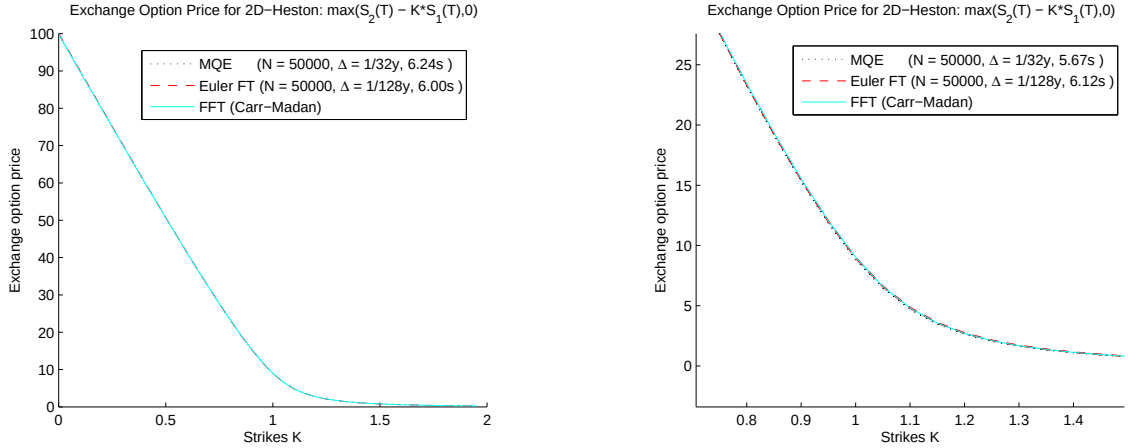


Figure 5.5: Exchange option price: this shows that MQE as well as Euler FT can attain high option price accuracy. The right-hand plot zooms in on the left-hand plot.

We plotted the semi-analytic option price once as a reference value. Just as in any of the coming test runs, five MQE prices and five Euler FT prices are plotted against this reference value. Each test run, we estimate the 95% confidence interval of both Monte Carlo prices from their respective paths. Furthermore, the average of the five CPU times of both Monte Carlo methods is displayed in the legend. In Figure 5.5, the three plotted functions are hard to distinguish since they all practically coincide. The confidence intervals show that both prices are indeed not significantly different from the semi-analytic price for all displayed strikes. Since both confidence intervals are small (less than 4% of the price estimate for all K), this shows the success of both numerical methods in this case.

³The computer used was an Intel(R) Core(TM)2 Duo CPU T9300 @ 2.50 GHz

We stated that $\rho_1 = 0$ is a necessary condition to be able to find a semi-analytic price. We illustrate this fact in Figure 5.6, a copy of the previous test, except that now we set $\rho_1 = -0.7$.

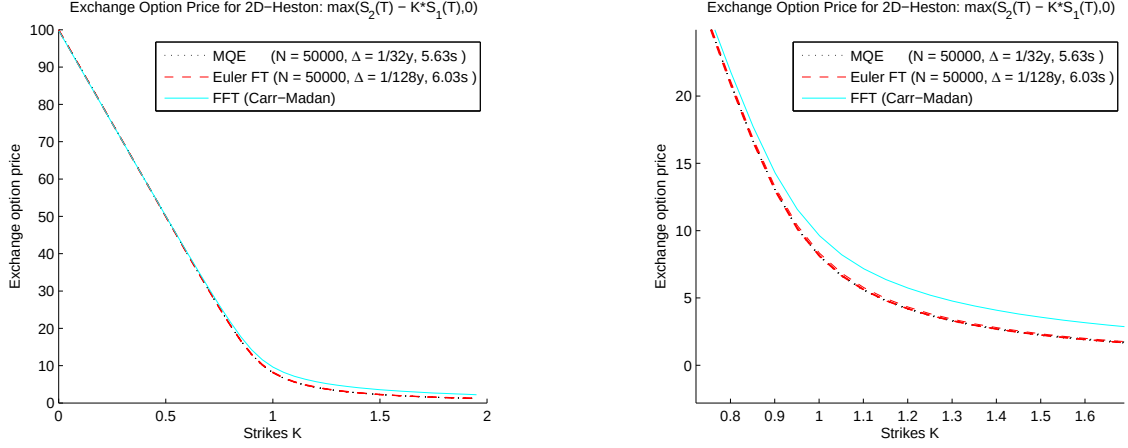


Figure 5.6: Exchange option price: this illustrates that we cannot find a semi-analytic option price for $\rho_1 \neq 0$. The right-hand plot zooms in on the left-hand plot.

In this figure, the same numbers of time steps and Monte Carlo paths are used as was done in the test run of Figure 5.5. The figure suggests that again the MQE prices and Euler FT prices mutually coincide. The fact that each estimate lies in the 95% confidence interval of the other estimate, supports this fact statistically. However, both Monte Carlo prices are significantly different from the semi-analytic price for $K \in [0.50, 1.95]$. This thus agrees with our derivation that $\rho_1 = 0$ is a necessary condition to find an accurate semi-analytic price.

We showed in Figure 5.5 that the Euler FT scheme can give option prices with sufficient accuracy too. One may wonder why we did not use the Euler FT scheme as a reference method in the first place. The answer is that we were not sure about this accuracy: as illustrated in Section 3.1.1, it is well-known that the Euler FT prices contain bias when the Feller condition is violated. How large this bias would be in case of multi-dimensional Heston, was just not known until we performed this test.

Moreover, Euler FT prices are not accurate for all parameter sets: consider for example a long maturity exchange option, with $T = 10$, and parameters:

	S_1	S_2
r	0	0
ε_i	1.5	1
κ_i	1.5	1.5
$\theta_i, V_i(0)$	0.09	0.04

Table 5.4: Parameter set of stocks S_1 and S_2

Note that we choose $b = 1.5$ this time. Suppose furthermore that the correlation structure is given by

$$\Sigma = \begin{pmatrix} 1 & \rho_{V_1 V_2} & \rho_1 & \rho_{V_1 S_2} \\ \rho_{V_1 V_2} & 1 & \rho_{V_2 S_1} & \rho_2 \\ \rho_1 & \rho_{V_2 S_1} & 1 & \rho_{12} \\ \rho_{V_1 S_2} & \rho_2 & \rho_{12} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & -0.5 \\ 1 & 1 & 0 & -0.5 \\ 0 & 0 & 1 & 0.8 \\ -0.5 & -0.5 & 0.8 & 1 \end{pmatrix}.$$

The computed exchange option values are given in Figure 5.7.

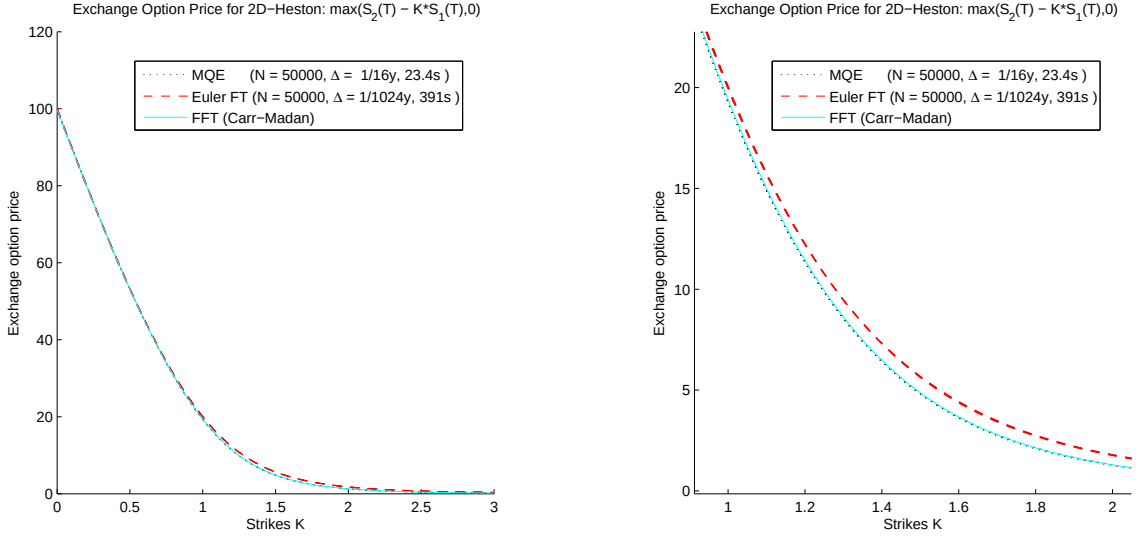


Figure 5.7: Exchange option price: this shows that the Euler FT scheme has significant bias for some parameter sets, even for large numbers of paths and steps. The right-hand plot zooms in on the left-hand plot.

Figure 5.7 suggests that all MQE prices again coincide with the semi-analytic price, whereas the Euler FT prices seem to contain bias. The 95% confidence intervals support this fact: the MQE prices are not significantly different from the semi-analytic price. The Euler FT prices exhibit significant bias for $K \in [0.75, 3]$. Further investigation shows that this bias is also noticeable for other parameter sets, as long as the maturity is large and the Feller condition is violated. We can explain this by the fact that for large maturity options on underlyings with high volatility of volatility parameter, the Euler FT scheme inherits too much bias during a too long period. Without the characteristic function, we might have concluded falsely that MQE inherits bias in case of long maturity. This shows the necessity of the test of this section.

How many times steps and paths are necessary to achieve a satisfactory accuracy? Let us assume the parameter set of the first test run of this section again (see Table 5.3). The right-hand plot in Figure 5.8 shows that 32 steps per year is too few for Euler FT to reach satisfactory accuracy. In fact, the Euler FT prices are significantly different from the semi-analytic value for $K \in [0.80, 1.95]$, whereas the MQE prices are not significantly different from the semi-analytic value for any strike. Note that the MQE and Euler FT CPU times are more less the same, which shows that MQE performs very well in this case.

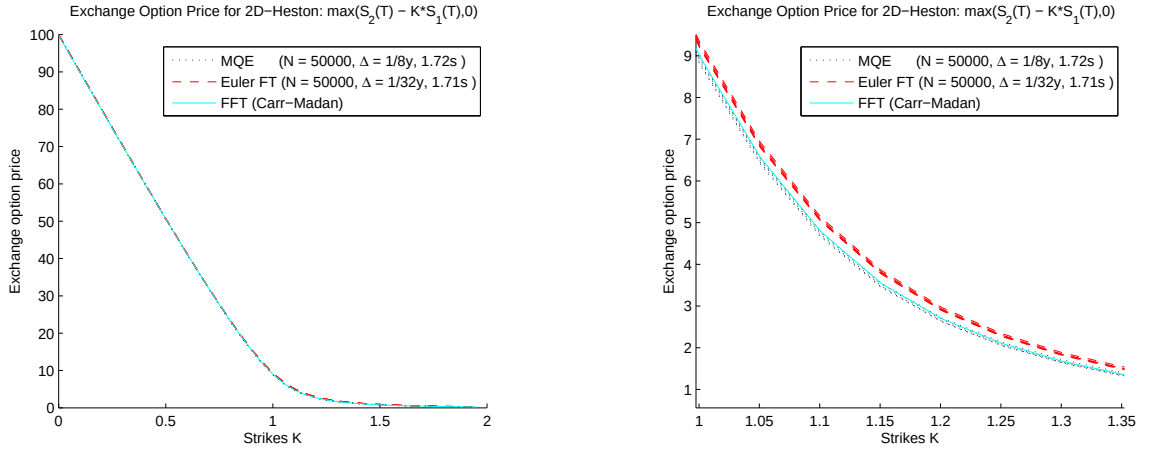


Figure 5.8: Exchange option price: this shows the superiority of MQE to Euler FT in accuracy of European call option pricing. The right-hand plot zooms in on the left-hand plot.

In Figure 5.9, the quality of the MQE scheme compared to the Euler FT scheme is illustrated. We used the practical time step of one quarter of a year for both Monte Carlo methods.

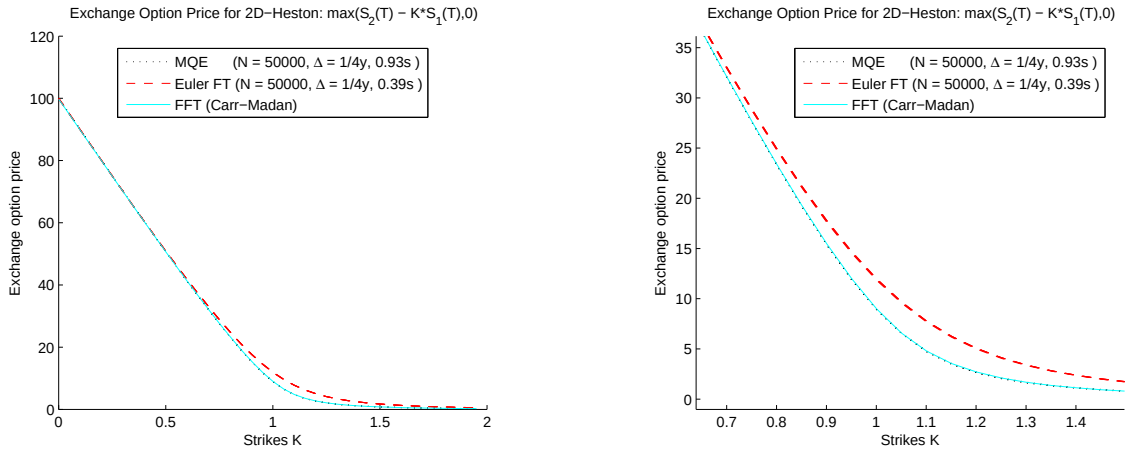


Figure 5.9: Exchange option price: this shows that MQE, unlike Euler FT, still has high accuracy in case of a large step size. The right-hand plot zooms in on the left-hand plot.

As the above figure suggests, the five MQE option prices are not significantly different from the semi-analytic value. The five Euler FT prices are structurally too large for at the money options: they differ significantly from the semi-analytic price for $K \in [0.50, 1.95]$.

5.5 Conclusion

The MQE method has passed several tests. The resulting stock-stock correlations and the corresponding stock-variance correlations after a computation are as prescribed beforehand. The MQE European call values are in accordance to the Euler FT prices when the Feller condition is satisfied. When this condition is violated, we can not use Euler FT as a reference, which is why we performed two other tests. The MQE scheme passed the test of validating a European call on a sum of three identical stocks in case the Feller condition was violated.

Since this test was somewhat trivial, we performed a second test: we priced an exchange option with MQE (and Euler FT), and reduced the dimension by a change of numéraire. In this way, we obtained a semi-analytic option price as a

reference value. One limitation represents the restrictions on the Heston parameter set: we had to choose zero correlation between the first stock process and the first variance process, and we had to choose the first variance process as a multiple of the second variance process.

MQE passed this test too: the MQE prices practically coincide with the semi-analytic price for sufficiently large number of Monte Carlo paths and time steps. When the Feller condition is violated, the Euler FT prices contain significant bias in case of long maturity options ($T = 10$), even for a very high number of steps and paths. In all other tests, using a sufficient number of Monte Carlo paths and steps, the Euler FT scheme could have served as a reference value for the MQE scheme.

Furthermore, the MQE prices are still very close to the semi-analytic price for a small number of time steps (for example, a time step of one quarter with a maturity of one year). The Euler FT prices contain large bias for this large time step. We already expected this advantage of MQE from the one-dimensional case.

We can conclude that the two-dimensional MQE scheme validates European call prices accurately and imposes correlation between stock processes correctly. From these facts, we have gained confidence that the MQE scheme performs well in two dimensions. Since more than two dimensions is a straightforward extension of the two-dimensional case, the MQE scheme will perform well for an arbitrary number of dimensions.

5.6 Further investigation

The MQE scheme can be extended by altering the correlation structure. We expect that dropping the assumption of zero correlation between $S_i(t)$ and $V_j(t)$, $i \neq j$ (the *cross-correlations*), will be most easy to implement. In this case, we could perform the construction of the MQE scheme in Section 4.2 as follows. The correlation matrix Σ will be different, since the submatrix Σ_{XV} in (4.5) will be a full matrix instead of a triangular matrix. This will alter the lower triangular matrix L , obtained by a Cholesky decomposition of $\Sigma = LL^T$. However, all derivations in that section can be made. They will lead to a form as in (4.12) with an extra term, containing a sum of other Itô integrals. These Itô integrals are of the form

$$\int_t^{t+\Delta} \sqrt{V_j(s)} d\tilde{W}_i(s), \quad i \neq j.$$

Now note that, given $\mathcal{V}_i$,

$$\int_t^{t+\Delta} \sqrt{V_j(s)} d\tilde{W}_i(s) \stackrel{d}{=} s_j Z_i \stackrel{d}{=} s_j \frac{f_i}{s_i}, \quad (5.23)$$

where we used the notations given at the end of Section 4.2 and assumed that $s_i \neq 0$. This last assumption is not valid for $\hat{V}_i(t) = \hat{V}_i(t + \Delta) = 0$ if we use the approximation

$$s_i^2 \approx \Delta[\gamma_1 \hat{V}_i(t) + \gamma_2 \hat{V}_i(t + \Delta)],$$

as proposed in (4.9). However, by choosing an approximation which can not be zero (for example, by using the first moment of s_i , which is derived in [Dufresne, 2001]), this problem can be avoided. In this way, we can sample these new Itô integrals exactly by use of representation (5.23). We can implement this extended expression just as we did with expression (4.12).

The other assumption, that we could avoid, would be the assumption of the zero variance-variance correlations. This would be more desirable than the avoidance of the previous assumption, since one naturally would expect dependence between the variances of stocks. When a group of stocks on average becomes more volatile, some traders will trade other stocks as well, which in turn will increase their volatility.

However, when constructing this version of the MQE method, one problem occurs. Since we use the QE method to compute the next step variance, the corresponding random variables are not normally distributed. Therefore, there is no exact method available to draw these random variables with the correct correlation structure. As mentioned in Section 4.1, one could nevertheless use the NorTA method to correlate these non-normally distributed random variables. In this way, the MQE extension can be constructed in the same way as we constructed the MQE extension with non-zero

cross-correlations. An investigation should be made of the error between the estimated variance-variance correlation and the originally imposed variance-variance correlation. If this error appears to be sufficiently small, we can adopt this MQE extension.

Chapter 6

Calibration of the MQE scheme

In finance, a calibration method is a method to choose values for the model parameters such that the model fits some chosen data well enough. The choice of parameter values is crucial for our MQE scheme to price derivatives accurately.

We can divide calibration methods in two main categories: historical data calibration and market price implied calibration. The former uses historical time series (mainly of stock prices) to estimate the current parameters. The latter uses the assumption of arbitrage absence to claim that derivatives with the concerned underlying are priced correctly in the market. The calibration method searches for parameter values such that the model derivative price is close enough to the market derivative price. More precisely, a market price implied calibration method minimizes the function

$$\sum_{i=1}^N w_i \left[C_i^{\bar{\Omega}}(K_i, T_i) - C_i^M(K_i, T_i) \right]^2 \quad (6.1)$$

over a discretized parameter space $\bar{\Omega}$. Here, N is the number of options used in the calibration, $C_i^{\bar{\Omega}}(K_i, T_i)$ and $C_i^M(K_i, T_i)$ are the respective model and market derivative prices, with strike K_i and maturity T_i , and w_i are weights. There are two main drawbacks of the historical data calibration method. Firstly, several parameters are not always observable in the market. Secondly, empirical estimates do often disagree with the market price implied estimates (see for example [Bakshi et al., 1997]). In that case, since we assume no arbitrage, we will prefer the market price implied estimates. Historical estimates will only be used if market price implied estimates are unavailable.

The curse of a more realistic model is the increase in complexity of the corresponding market implied calibration method. This also holds for the Heston model: it is able to capture the volatility smile, whereas the Black-Scholes model is not since the volatility is then assumed constant. However, in the Black-Scholes model, one only needs to calibrate the volatility parameter¹, whereas in the one-dimensional Heston model, five parameter values need to be found ($\kappa, \theta, \varepsilon, V(0)$ and ρ). It is clear that the computational effort will increase much when the parameter space of the optimization method is five-dimensional instead of one-dimensional.

The situation is similar in case of a D -dimensional Heston process, compared to D independent Heston processes. We assume the same correlation structure as in the MQE scheme. Our model contains $5D$ single-asset Heston parameters plus $D(D-1)/2$ stock-stock correlation parameters. The model consisting of D independent Heston submodels can be calibrated in D separate steps, calibrating five Heston parameters each step. Therefore, we decide to divide the calibration method in two steps:

1. We calibrate the five parameters of each single-asset Heston submodel: $\kappa_i, \theta_i, \varepsilon_i, V_i(0)$ and ρ_i , $\forall i = 1, \dots, D$. Therefore, we use a known Heston calibration method.
2. We calibrate the remaining parameters, i.e. the correlations between the Heston submodels.

¹Assuming no arbitrage, the drift of the each stock is equal to the risk-free rate. Its calibration can thus be performed independent from the considered model calibration, and is beyond the scope of this thesis.

By assumption of our correlation structure, the second step involves the calibration of the stock-stock correlations only. Note that the corresponding stock-variance correlation, ρ_i , will be calibrated in step 1 already. This separation of the calibration of Σ will not yield any risk of positive indefiniteness. All variance processes are only correlated with their corresponding stock process. Therefore, Σ will be positive definite as long as the stock-stock correlation matrix, Σ_X , is positive definite.

We will treat the two steps in the coming two sections, respectively. In the third section, we will construct a market implied calibration method for the stock-stock correlations. In the fourth and fifth section, the method will be investigated and discussed. The last section contains a conclusion on the method as well as recommendations for further investigation.

6.1 Single-asset Heston calibration

The first step of our multi-dimensional calibration method has been developed as a calibration technique for derivatives with one underlying. Therefore, it has been investigated extensively. See for example [Moodley, 2005] and [Bin, 2007]. We will give an overview on the considerations of a one-dimensional Heston calibration method.

As we explained, the main problem of a single-asset Heston calibration is the dimensionality of the optimization. One solution for this problem is to narrow the parameter space by insights on the derivative of concern. The sensitivity of the derivative price to a certain Heston parameter can vary much per Heston parameter. We will explain this by giving an interpretation of every parameter. Since we consider the single-asset Heston model, we will omit the indices of the Heston parameter, just as in Chapter 2 and 3. The long-term variance, θ , represents the mean value of the variance process. A first guess for this parameter could be chosen as the square of the implied volatility obtained by a Black-Scholes calibration. Then one could concentrate the optimization grid for this parameter around this first guess. The initial variance, $V(0)$, influences the option price significantly when the maturity is short. A first guess for this parameter value could be the volatility implied by a Black-Scholes calibration of an at the money option with short maturity. The volatility of volatility, ϵ , affects the kurtosis of the stock distribution. High volatility of volatility implies heavy tails on both side of the distribution. In contrast, the mean reverting term, κ , drives the variance process back to the long term variance. The *volatility surface* may employ a first guess of these two parameters. This surface consists of volatility estimates implied by Black-Scholes calibrations for different strikes and maturities.

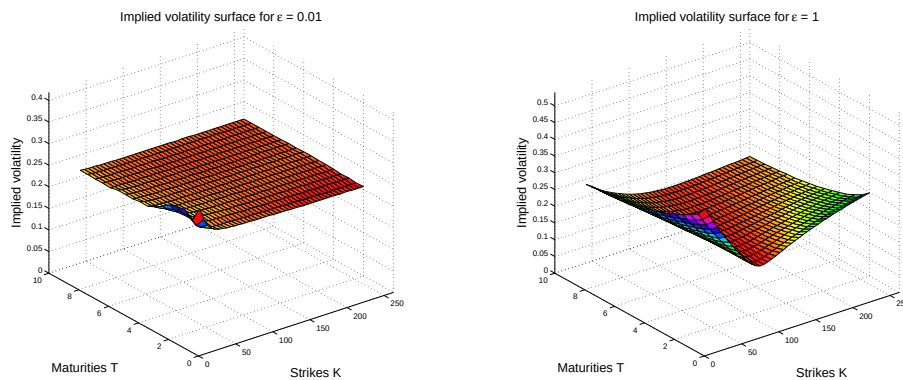


Figure 6.1: Implied volatility surfaces for low and high value of ϵ . $\kappa = 1.5, \theta = V(0) = 0.04, \rho = -0.6$ in both cases.

The implied volatility surface will exhibit a smile for high ϵ , as can be seen in Figure 6.1. As the mean reverting term, κ , drives the variance process back to its mean, a high value of κ will moderate the smile. Therefore, a first guess of these two parameters can be based on the shape of the volatility surface.

The stock-variance correlation, ρ , affects the skewness of the stock distribution. Note that for negative ρ , a volatil-

ity increase will stimulate the stock value to decrease. Another way of saying this is that the left tail of the stock distribution will be heavier than the right tail, thus inducing a more negative skew, as can be seen in Figure 6.2.

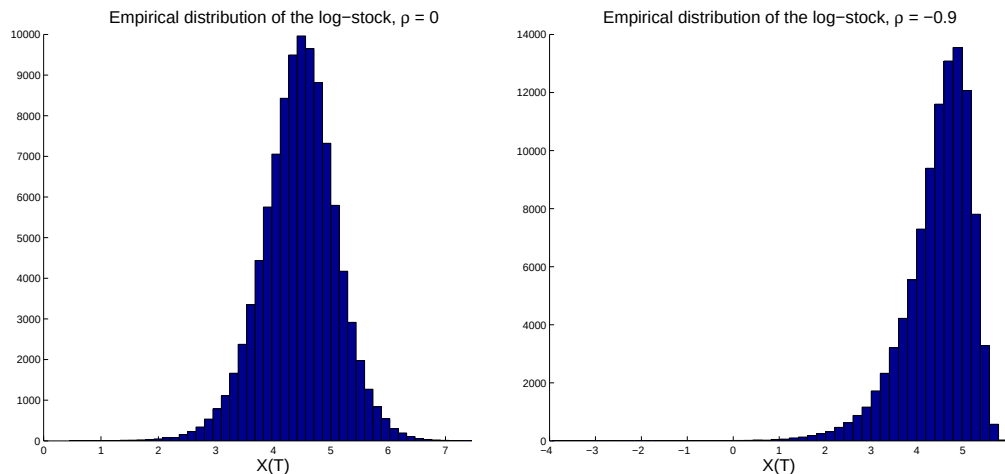


Figure 6.2: The influence of the stock-variance correlation on the distribution of the log-stock. The right-hand plot exhibits a negative skew. Further parameters: $\kappa = 1.5, \theta = V(0) = 0.04, \varepsilon = 0.5$.

Options which are rather sensitive to the skew (like cliquets², see [Dupire, 1994]), can be calibrated more efficiently by putting educated restrictions on the optimization grid of ρ . For example, one may restrict the correlation to the interval $[-1, 0]$. The motivation is the expectation that when a stock becomes more volatile, its value often decreases.

We conclude that the optimization parameter space should be anticipated on the parameter to which the concerned derivative is sensitive. By restricting the optimization parameter space, a problem may occur, however. For two different derivatives with the same underlying, one may obtain two different optimal parameter sets. For the validation of a financial product which is based on both derivatives, it is unclear which of the parameter sets should be used. One should avoid this problem by combining the restrictions of all different subproducts in the calibration.

The anticipation of the optimization algorithm can be done in several ways. For example, one could first calibrate all parameters except the stock-variance correlation, by use of variance swap prices. The payoff of this option is based on the realised variance of an asset, which is why one could price it using the Heston variance process only. After that, one can calibrate the stock-variance correlation separately, thus emphasizing the optimization on this parameter.

One drawback of this procedure is that variance swaps are not known to be very liquid. Therefore, prices can be unrealistic or even missing for certain extreme maturities and strikes. One could use an extrapolation method to substitute missing quotes with estimates. However, the derivative price will depend on the used extrapolation method in that case, which is undesirable. Instead, one could anticipate the optimization algorithm by fixing one or two less important parameters so that the optimization method can search the more important parameters from relatively fine grids. Fixing can be based on the mentioned first guesses we gave at the start of this section.

The objective function of the minimization problem in (6.1) is not linear in case of a single-asset Heston model. Further, it is often far from being convex and there are often many local minima (see [Mikhailov & Nögel, 2003]). Therefore, several optimization procedures have been investigated, which can be divided in two main categories:

Local algorithms

One chooses an initial guess for the parameter set. Depending on the gradient of the objective function in that point, an optimal direction is estimated. Following that gradient ('going downhill'), one arrives at a new point, and estimates

²A cliquet option is an option consisting of a series of consecutive forward start options. The first is active immediately. The second becomes active when the first expires, and so on. Each option is struck at-the-money when it becomes active

a new optimal direction again. This iterates until a minimum is reached. One always has the risk of arriving at a local minimum, which makes the initial guess crucial.

Stochastic algorithms

In order to avoid arriving at a local minimum, one can include stochastic jumps in the algorithm. The algorithm searches downhill, but may search uphill with a probability decreasing in time. There are theorems that state that the algorithm always arrives at the global minimum, provided that the probability decrease is sufficiently slow (see [Kirkpatrick, 1984]). These stochastic algorithms are computationally more expensive than the local algorithms.

6.2 Stock-stock correlation calibration

Stock-stock correlations can have significant influence on derivative prices. In Figure 6.3, a small example is shown of the sensitivity of a European call price to ρ_{12} :

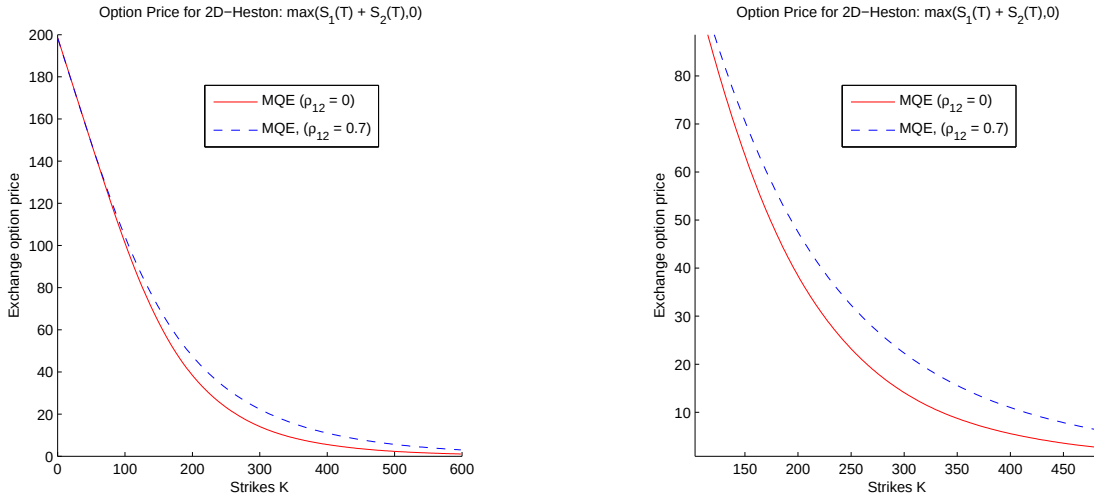


Figure 6.3: Sensitivity of call price to stock-stock correlation ρ_{12} . Other parameter values: $S_1(0) = S_2(0) = 100$, $\kappa_1 = 1$, $\kappa_2 = 1.5$, $\theta_1 = V_1(0) = 0.04$, $\theta_2 = V_2(0) = 0.05$, $\varepsilon_1 = 0.01$, $\varepsilon_2 = 0.02$, $\rho_1 = -0.4$, $\rho_2 = -0.5$, $T = 10$, $\Delta = 0.125$ year, number of Monte Carlo paths is 10.000. The right-hand plot zooms in on the left-hand plot.

In the above figure, the prices of two European call options on the sum of two stocks have been estimated by use of the MQE method. The difference between these two Monte Carlo runs is the value of ρ_{12} . All other parameters coincide, as well as the Brownian motions samples. One can see that for most strikes, a change in correlation causes a change in the option price. In order to state this more precisely, we estimated a 95% confidence interval from the Monte Carlo paths of both option prices. For $K \in [119, 600]$, these two intervals are disjoint, which means that the option prices differ significantly for these strikes. This illustrates the necessity of an accurate calibration method. We found one historical calibration method in literature, and we will discuss this one first.

6.2.1 Literature on multi-asset Heston calibration

The authors of [Dimitroff et al., 2009] agree with us to calibrate the one-dimensional models separately by known methods first. They assume the same correlation structure as used in the MQE scheme. As a next step, they calibrate each stock-stock correlation separately, by use of the historical correlations of the log-stock returns:

$$\hat{\rho}_{ij}^R := \text{corr}(\hat{X}_i(t + \Delta) - \hat{X}_i(t), \hat{X}_j(t + \Delta) - \hat{X}_j(t)),$$

with $\hat{X}_i(t)$ the observed value of the log-stock $X_i(t)$ in the market at time t . We assume that $\hat{\rho}_{ij}^R$ is the realization of the theoretical log-stock return correlation:

$$\rho_{ij}^R := \text{corr}(X_i(t + \Delta) - X_i(t), X_j(t + \Delta) - X_j(t)).$$

The article reasons that unlike in the Black-Scholes model, $\rho_{ij}^R \neq \rho_{ij}$, with ρ_{ij} the correlation between the corresponding stock Brownian motions. In case of Black-Scholes dynamics, equality *does* hold since the volatility is deterministic (see (5.1), which is obviously not true in case of Heston dynamics. Therefore, one has to find a stock-stock correlation matrix so that the resulting correlation between the log-stock returns will be close to the historical correlations. More strictly, one has to find:

$$\min_{\Sigma_X \in C_D} \left\| \Sigma_R(\Sigma_X) - \hat{\Sigma}_R \right\|, \quad (6.2)$$

with C_D the set of all correlation matrices of dimension D and the matrices defined by

$$\Sigma_R = (\rho_{ij}^R)_{1 \leq i, j \leq D}, \quad \Sigma_X = (\rho_{ij})_{1 \leq i, j \leq D}, \quad \hat{\Sigma}_R = (\hat{\rho}_{ij}^R)_{1 \leq i, j \leq D}.$$

The method explains how to solve this minimization problem per element of Σ_X :

$$\min_{-1 \leq \rho_{ij} \leq 1} |\rho_{ij}^R(\rho_{ij}) - \hat{\rho}_{ij}^R|. \quad (6.3)$$

Firstly, a limiting relationship is found between ρ_{ij}^R and ρ_{ij} as the step size goes to zero: ρ_{ij}^R is approximately linear in ρ_{ij} , with a positive slope. This insight gives rise to use a line search method (for example, the bisection method) to solve (6.3). Throughout this procedure, $\rho_{ij}^R(\rho_{ij})$ is required, which is estimated by using a small Monte Carlo method: sample paths are generated by using Brownian motions with correlation ρ_{ij} . Then the sample path average of the correlation between the log-stock returns is the estimate for $\rho_{ij}^R(\rho_{ij})$. The authors expect that the resulting matrix Σ_X^* with optimal elements ρ_{ij}^* will be close to the solution of (6.2). However, it could be that Σ_X^* is not positive definite. To transform Σ^* to a ‘close by’ positive definite matrix in this case, the article recommends the procedure described in [Jäckel, 2002].

This calibration method has some drawbacks. Firstly, one has to run a Monte Carlo method at each optimization step in the optimization of each ρ_{ij} , which will be computationally expensive. Secondly, the resulting correlation matrix might differ much from the solution of (6.2), since the minimization is performed per element of Σ_X . Moreover, the method is a historical calibration method, and we explained at the start of this chapter that market price implied calibration method are preferred.

6.3 Market implied calibration method

Since market implied calibration is preferred to historical calibration, we desire an alternative calibration method of the former category. We assume that each one-dimensional submodel has been calibrated separately. Now, we require a fast method to compute market prices given a chosen parameter set of stock-stock correlations. Then one can efficiently search for the optimum in a high-dimensional parameter space.

In the one-dimensional case, a common procedure is to use a Fourier method to compute European call and put prices, since the characteristic function of the log-stock value is available. In the multi-dimensional extension, one may try to calibrate in a similar way, namely via quoted call (or put) prices on a *basket option*. Its payoff is given by

$$\max \left[\sum_{i=1}^D S_i(T) - K, 0 \right]. \quad (6.4)$$

Unfortunately, the distribution or characteristic function of

$$S_s(t) := \sum_{i=1}^D S_i(t) \quad (6.5)$$

is not known explicitly for $D \geq 2$. We conclude that we can not find semi-analytic prices for this option in this way.

Since neither the characteristic function nor the density function of $S_s(T)$ is known, we need another analytic property depending on the stock-stock correlations. Therefore, we propose method based on moment matching. The *matching moment* technique is based on the substitution of a distribution which is mathematically more tractable, and which approximates the original distribution in the sense that a certain number of moments coincides. For example, the QE scheme uses two moment matching techniques. It approximates the noncentral chi-square distribution, $f_{\chi^2/2}$, by two other distributions, by setting their first two moments equal to the first two moments of $f_{\chi^2/2}$ (see Section 3.2.2). More of the topic of moment matching calibration techniques in finance can be found in [Levy, 1992] and [Turnbull & Wakeman, 1991].

The idea of moment matching originates from the fact that moments of $S_s(T)$ higher than the first moment, depend on all stock-stock correlations. For example, the second moment of $S_s(T)$ can be written as:

$$\mathbb{E}[S_s^2(T)] = e^{2 \int_0^T r(t) dt} \sum_{i=1}^D \sum_{j=1}^D S_i(0) S_j(0) \mathbb{E} \left[\exp \left\{ \int_0^T \sqrt{V_i(t)} \rho_{ij} \sqrt{V_j(t)} dt \right\} \right]. \quad (6.6)$$

One can find the derivation in Appendix B. Now suppose the following:

1. Suppose that we find a close approximation, $\hat{M}_2^{S_s}(T)$, for $\mathbb{E}[S_s^2(T)]$.
2. Suppose that we find a close approximate distribution, F , for the distribution of $S_s(T)$, of which we can derive the second moment, $M_2^F(T)$. Suppose further that we can calibrate the corresponding approximate model sufficiently fast (for example, by existence of an analytic solution, or by use of a Fourier method).

Then, our proposed market implied calibration method will be as follows. We calibrate our model to given market prices, assuming $S_s(T) \sim F$, and find its second moments, $M_2^F(T)$, as function of the maturity T . This is our observed data set. We choose Σ_X so that the two moments $\hat{M}_2^{S_s}(T)$ and $M_2^F(T)$ are close for all T . In this way, we have calibrated all stock-stock correlations. The two above mentioned suppositions are discussed in the respective next two sections.

6.3.1 Approximate second moment of S_s

As it is obvious from (6.6), the derivation of an analytic expression of $M_2^{S_s}$ requires the derivation of

$$E_{ij} := \mathbb{E} \left[\exp \left\{ \rho_{ij} \int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \right\} \right]. \quad (6.7)$$

We did not found any analytic expression for E_{ij} in literature, and we expect that it is hard to find one. Instead, we propose some estimates, each based on one of the following substitutions, with $m = i, j$:

1. For all t , we use the estimate

$$V_m(t) \approx \theta_m. \quad (6.8)$$

This is a strong assumption, we expect bad behavior when the Feller condition is violated. We will refer to the resulting estimate of E_{ij} by $\hat{E}_{ij}^\theta$.

2. $V_m(t) \approx \mathbb{E}[\sqrt{V_m(T)}]$, where we use the expression from [Grzelak & Oosterlee, 2010] for the right-hand side:

$$\mathbb{E}[\sqrt{V_m(T)}] = \sqrt{2c(t)} e^{-\lambda(t)/2} \sum_{k=0}^{\infty} \frac{1}{k} (\lambda(t)/2)^k \frac{\Gamma\left(\frac{1+d}{2} + k\right)}{\Gamma\left(\frac{d}{2} + k\right)},$$

with $\Gamma(t)$ the Gamma function and

$$c(t) = \frac{\varepsilon^2}{4\kappa} (1 - e^{-\kappa t}), \quad d = \frac{4\kappa\theta}{\varepsilon^2}, \quad \lambda(t) = \frac{4\kappa V_m(0) e^{-\kappa t}}{\varepsilon^2 (1 - e^{-\kappa t})}.$$

The estimation is based on a truncation of the infinite sum. One drawback is that the computation of the Gamma function becomes expensive for small values of ε .

3. $V_m(t) \approx \mathbb{E}[\sqrt{V_m(T)}]$, but now we use the following estimate as proposed in [Grzelak & Oosterlee, 2010]:

$$\mathbb{E}[\sqrt{V_m(T)}] \approx \sqrt{c(t)(\lambda(t) - 1) + c(t)d + \frac{c(t)d}{2(d + \lambda(t))}}.$$

The authors of [Grzelak & Oosterlee, 2010] show that this estimate may not be well-defined when the Feller condition is violated. They propose to use the exact representation in this case.

4. $\int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \approx \sqrt{\mu_i \mu_j}$, with

$$\mu_m := \mathbb{E} \left[\int_0^T V_m(t) dt \right] = \theta_m T + (e^{-\kappa_m T} - 1)(\theta_m - V_m(0)) / \kappa_m.$$

The expression for this expectation is derived in [Dufresne, 2001].

For simplicity, we will initially use $\hat{E}_{ij}^\theta$ as our estimate for E_{ij} .

6.3.2 Approximate distribution of $S_s(T)$

The first candidate for F , the approximate distribution of $S_s(T)$, is the normal distribution: $F =: BS$. In fact, this approximation means that we assume that $S_s(t)$ approximately follows one-dimensional Black-Scholes dynamics. This suggests:

$$S_s(T) \approx S_s(0)e^{(r - \sigma^2/2)T + \sigma W(T)},$$

with constant risk-free rate r , constant volatility σ , and $W(t)$ a Brownian motion. This assumption implies that for Z an independent standard normal random variable,

$$\begin{aligned} M_2^{BS}(T) &= \mathbb{E}[S_s^2(0) \exp\{2rT - \sigma^2 T + 2\sigma W(T)\}] \\ &= S_s^2(0) e^{2rT - \sigma^2 T} \mathbb{E}[\exp\{2\sigma \sqrt{T} Z\}] \\ &= S_s^2(0) e^{2rT - \sigma^2 T + 2\sigma^2 T} \\ &= S_s^2(0) e^{(2r + \sigma^2)T}. \end{aligned}$$

The second candidate for F is the one-dimensional Heston distribution: $F =: HES$. Consider the joint characteristic function

$$\phi(u, v) = \mathbb{E}[e^{iuV(T) + ivx(T)}],$$

where $x(T) = \ln S(T) - \ln S(0)$, and $S(t)$ and $V(t)$ follow single-asset Heston dynamics. Its explicit form can be found in Appendix A of [Andersen, 2007], for example. Note that

$$\phi(0, -ni) = \mathbb{E}[e^{n(\ln S(T) - \ln S(0))}] = \frac{\mathbb{E}[S^n(T)]}{S^n(0)},$$

for $n = 1, 2, \dots$, which yields

$$M_2^{HES}(T) = S^2(0) \phi(0, -2i).$$

6.4 Investigation of the calibration method

The approximate distribution F must be close enough in the sense that their second moment must be close to the second moment of $S_s(T)$. In order to choose from these two candidate distributions, we will investigate the distribution of $S_s(T)$. Therefore, we will sample $S_s(T)$ with our MQE method and investigate the corresponding empirical distribution. We consider different number of underlyings ($D = 2$ and $D = 10$) and cases in which the Feller condition is satisfied and violated. For the sake of generality, we choose different Heston parameter sets per stock. For each stock in fact, we draw the Heston parameters uniformly from practical intervals, as shown in Table 6.1.

parameter	(a,b) such that parameter value $\sim U(a,b)$
$S_i(0)$	(10, 110)
κ_i	(0.5, 1.5)
$\theta_i = V_i(0)$	(0.02, 0.06)
ρ_i	(-1,0)

Table 6.1: Parameter distribution of all underlying stocks

The risk-free rate $r(t)$ is taken to be zero and a positive definite Σ_X is some arbitrary correlation matrix (in [Bendel & Mickey, 1978], a method is proposed to draw correlation matrices with uniformly distributed eigenvalues). Finally, we choose all volatilities of volatility from the same uniform distribution, considering three cases. The domain of the distribution will depend on the fulfillment of the Feller condition of all one-dimensional Heston models:

Case	<i>Feller Satisfied</i>	<i>Feller Almost Violated</i>	<i>Feller Violated</i>
(a,b) , so that $\varepsilon_i \sim U(a,b)$	(0.008, 0.01)	(0.08, 0.1)	(0.8, 1)

Table 6.2: Uniform distribution intervals for ε_i , for all i , for three different cases

Histograms of the empirical distribution of $\ln S_s(T)$ are displayed in Figures 6.4 and 6.5, considering 2 and 10 stocks, respectively. In both figures, the three histograms correspond to the three cases mentioned in Table 6.2.

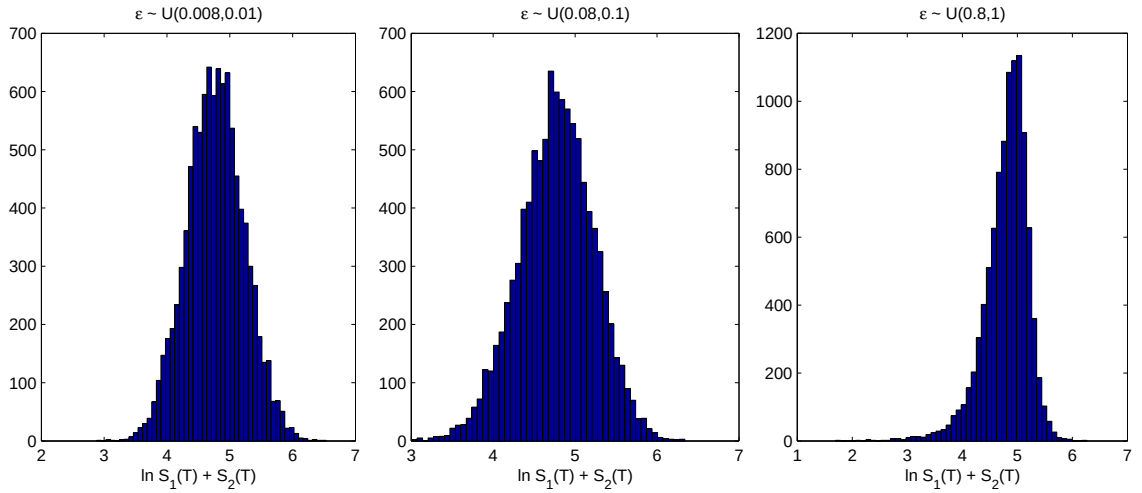


Figure 6.4: Sample histograms of $\ln S_s(T)$ for $D = 2$. From left to right: cases *Feller Satisfied*, *Feller Almost Violated* and *Feller Violated*.

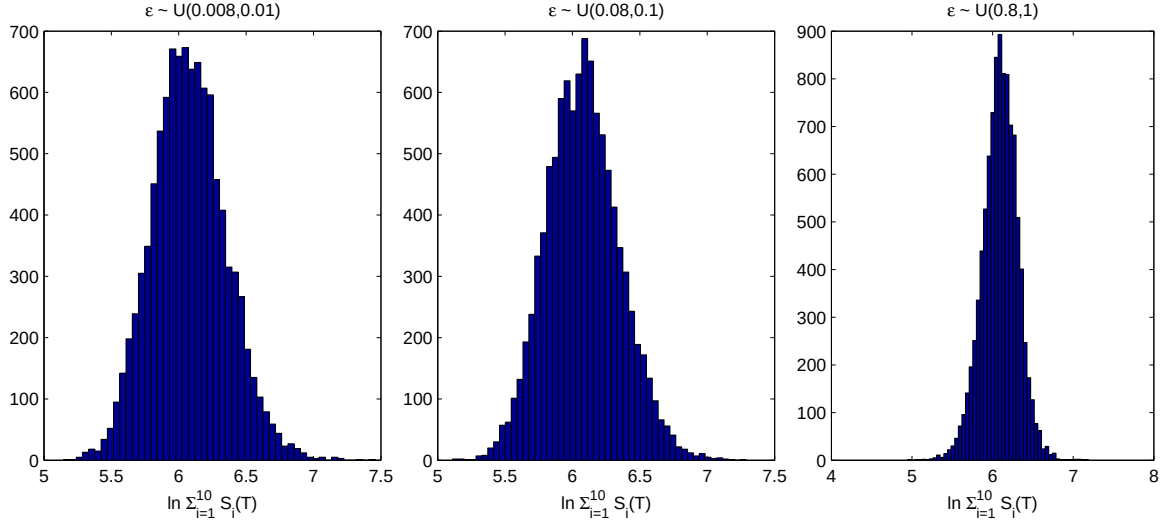


Figure 6.5: Sample histograms of $\ln S_s(T)$ for $D = 10$. From left to right: cases *Feller Satisfied*, *Feller Almost Violated* and *Feller Violated*.

One can see in Figures 6.4 and 6.5 that for small ϵ_i , the empirical distribution of $\ln S_s(T)$ is close to normal. For large ϵ_i however, this is not the case. It is hard to see from these histograms whether a Heston distribution will fit this data better. Therefore, we choose either $F = BS$ or $F = HES$, and perform the following test³:

1. We consider two stocks. We choose each one-dimensional Heston parameter set from the parameter distributions displayed in Tables 6.1 and 6.2. We choose some arbitrary stock-stock correlation (in our case $\rho_{12} = 0.5$). We let the MQE method simulate basket option prices with the payoff as given in (6.4), for different strikes and maturities. We assume that these prices are market quotes.
2. We calibrate these prices by assuming that $S_s(t) \sim F$. A plot of the simulated quotes against the calibrated prices is displayed in Figure 6.6.
3. We compute $M_2^F(T)$ for all considered maturities T . We also estimate the second moment from our MQE Monte Carlo paths. This estimate, $M_2^{MQE}(T)$, serves as a reference value. We expect that this estimate will be sufficiently accurate by taking a large number of Monte Carlo paths (1.000.000) and time steps ($\Delta = 0.1$ year). See Table 6.3.
4. We approximate $\mathbb{E}[S_s^2(T)]$ by substituting $E_{ij} \approx \hat{E}_{ij}^0$, using (6.8). We match this moment to $M_2^F(T)$, thus obtaining an optimal stock-stock correlation. We also estimate E_{ij} from our Monte Carlo Paths and match the thereby induced moment to M_2^{MQE} , thus obtaining a Monte Carlo reference optimum.
5. We investigate whether the optimal stock-stock correlation is close to the imposed ρ_{12} .

For the Black-Scholes calibration in step 2, the method uses the analytic Black-Scholes formula for a European call option price:

$$C(S_s(0), T) = S_s(0)\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$

with

$$\begin{aligned} d_1 &= \frac{\ln S_s(0) - \ln K + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \\ d_2 &= d_1 - \sigma\sqrt{T}. \end{aligned}$$

³The computer used was an Intel(R) Core(TM)2 Duo CPU T9300 @ 2.50 GHz

Here $C(S_s(0), T)$ denotes the call option value at time T , assuming initial stock price $S_s(0)$, $\Phi(\cdot)$ the standard normal cumulative distribution function. For each subsequent volatility, chosen from a certain grid, the induced option price is compared to the market quotes for every maturity and strike. The mean error of these Black-Scholes prices is the function that we will numerically minimize over the volatility parameter σ . For the Heston calibration, the method is similar. Now we use the COS method as described in [Fang & Oosterlee, 2008], to obtain the semi-analytic call price instead of the Black-Scholes call price. To find the optimal parameter set, we use a local minimizer, as mentioned in Section 6.1. We set the interval for the stock-variance correlation, ρ_i , to $[-1, 0]$. Further, we focus the long-term variance grid on the implied volatility obtained by the Black-Scholes calibration. In this way, we increase the efficiency of the Heston calibration method.

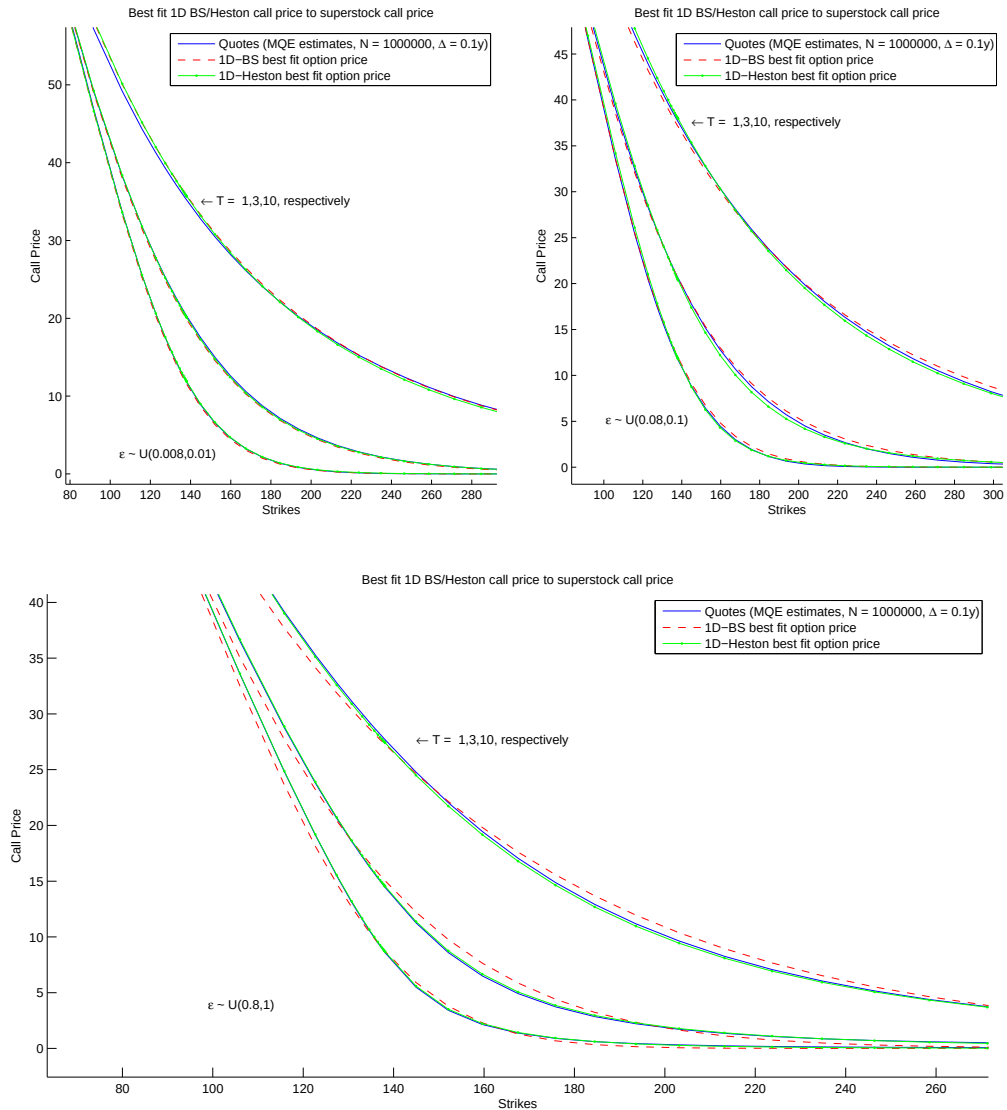


Figure 6.6: Performance of approximate distribution in fitting European call prices. All three figures contain simulated market quotes of the basket option, plotted against the best fits of these prices obtained by a one-dimensional Black-Scholes calibration and a one-dimensional Heston calibration. All for different strikes and maturities, and the parameters as given in Tables 6.1 and 6.2. The top left plot represents the case *Feller Satisfied*, the top right *Feller Almost Violated*, and the last plot *Feller Violated*

The Black-Scholes calibration took 3.2 seconds, while the Heston calibration took 456 seconds. At first sight, Figure

6.6 suggests that both the Black-Scholes distribution and Heston distribution fit the market prices well. In case Feller Almost Violated and case Feller Violated, the Black-Scholes calibration prices are slightly different from the market quotes. For the moment, this fact does not worry us much, since the quality of the approximation depends on the quality of the moment estimate only. In Table 6.3, the second moment estimates of $S_s(T)$ are displayed.

$M_2^F(T)$	<i>Feller Satisfied</i>			<i>Feller Almost Violated</i>			<i>Feller Violated</i>		
	$T = 1$	$T = 3$	$T = 10$	$T = 1$	$T = 3$	$T = 10$	$T = 1$	$T = 3$	$T = 10$
$M_2^{BS}(T)$	19229	20226	24141	19210	20169	23914	18977	19442	21160
$M_2^{HES}(T)$	19234	20237	24166	19222	20181	23911	19072	19663	21780
$M_2^{MQE}(T)$	19220	20244	24204	19216	20174	23946	19075	19766	22129
$se_{MQE}(T)$	6.3	12.1	33.2	6.1	11.7	31.1	5.2	140.7	224.7

Table 6.3: $M_2^F(T)$, as estimated by using two calibration methods and the MQE Monte Carlo estimate. The last row displays the standard error of $M_2^{MQE}(T)$.

One can see in Table 6.3 that the standard error of $M_2^{MQE}(T)$ is small except for large ε and T . This indicates that $M_2^{MQE}(T)$ is often an accurate estimate for the second moment of $S_s(T)$. The values of $M_2^{BS}(T)$ and $M_2^{HES}(T)$ differ only a few standard errors from $M_2^{MQE}(T)$ in all cases. This suggests that these moment estimates are accurate too. One can see that $M_2^{HES}(T)$ is closer to the MQE estimate than $M_2^{BS}(T)$ for large ε , which is intuitive since the Heston dynamics are a generalization of the Black-Scholes dynamics.

The calibrated stock-stock correlations obtained in this test, are displayed in Table 6.4.

$M_2^F(T)$	<i>Feller Satisfied</i>	<i>Feller Almost Violated</i>	<i>Feller Violated</i>
$M_2^{BS}(T)$	0.46	0.40	-0.48
$M_2^{HES}(T)$	0.47	0.40	-0.26
$M_2^{MQE}(T)$	0.46	0.39	-1.00

Table 6.4: Calibrated stock-stock correlation, as obtained by using three different approximations for S_s , and for three different levels of the volatility of volatility. The imposed correlation is $\rho_{12} = 0.5$.

Table 6.4 shows that our method performs quite well in case all volatility of volatility parameter values are low. Let us define the calibration error as the absolute difference between ρ_{12} and the calibrated estimate of ρ_{12} . Further investigation shows that the calibration error increases as ε_1 and ε_2 increase. Since the CPU time of one calibration run is around 30 minutes, we did not collect statistics of the error mean over multiple runs. Nevertheless, several calibration runs show that the calibration error is less than 0.05 in case *Feller Satisfied* and less than 0.13 in case *Feller Almost Violated*. One may think that the calibration method is successful as long as the volatilities of volatility of the two underlyings are sufficiently small. However, one should be aware of the sensitivity of the derivative to stock-stock correlation. In case of high sensitivity (like is the case with dispersion trades), the performance shown under Feller Almost Violated will certainly not be sufficient to price such derivative accurately. In case *Feller Violated*, the calibration is obviously too large for any practical purpose.

6.5 Discussion of the calibration method

This error is partly caused by the estimation error of (6.8). As we already explained in Section 6.3.1, this latter error will increase in ε_m , which is reflected in the above numerical results. However, we do not notice better calibration performance for any other choice of the four proposed approximations for E_{ij} . Furthermore, the calibration method performs bad too when using MQE estimates in case *Feller Violated* (see Table 6.4). This indicates that there must be another problem, which we revealed after a further investigation. Note that all four estimates of E_{ij} are based on the following approximation:

$$E_{ij} = \mathbb{E} \left[\exp \left\{ \rho_{ij} \int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \right\} \right] \approx \exp \left\{ \rho_{ij} \mathbb{E} \left[\int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \right] \right\}.$$

For large ε_i and ε_j , the above integral will have a large standard deviation, since the processes $V_i(t)$ and $V_j(t)$ will be more volatile. Taking the exponential function of this integral will increase the standard deviation of the approximate distribution of $S_s(T)$ exponentially. When estimating E_{ij} numerically, an increase of the volatility of volatility parameter will significantly decrease the accuracy of the estimate. In fact, further investigation of the Monte Carlo paths shows that the standard error of the E_{ij} Monte Carlo estimate is often around 20% of the estimate itself. Since the value of the second moment is rather insensitive to changes in ρ_{ij} , we need an accurate estimate. We conclude that the standard error is too large to achieve this. This explains why the performance of the calibration method based on MQE estimates, decreases in ε_1 and ε_2 .

A significant improvement of the approximation of E_{ij} will be the key to a robust calibration method. To this purpose, note that E_{ij} simplifies for $i = j$. Since $\rho_{ii} = 1$ and $V_i(t) \geq 0$ for all t , one can write

$$E_{ii} := \mathbb{E} \left[\exp \left\{ \int_0^T V_i(t) dt \right\} \right].$$

In [Dufresne, 2001], the moment generating function of the integrated square-root process is derived:

$$MGF(s) := \mathbb{E} \left[\exp \left\{ -s \int_0^T V_i(t) dt \right\} \right], \quad s \geq 0.$$

It is tempting to set $E_{ii} = MGF(-1)$. Unfortunately, this equality does not hold since the moment generating function has positive domain only. Numerical experiments show that this estimate is indeed inaccurate, unless the volatility of volatility parameter is very small. Our last attempt for a close approximation of E_{ij} involves the power series of the exponential function:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This expansion suggests the following approximation

$$E_{ij} = \mathbb{E} \left[\exp \left\{ \rho_{ij} \int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \right\} \right] \approx 1 + \sum_{n=1}^N \frac{\rho_{ij}^n}{n!} \mathbb{E} \left[\left(\int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt \right)^n \right],$$

with $N \in \mathbb{N}$ not too large. The estimate will therefore be a weighted sum of the first N moments of the integral

$$\int_0^T \sqrt{V_i(t)} \sqrt{V_j(t)} dt.$$

For $i = j$, these moments can be derived analytically by use of the moment generating function $MGF(s)$. The first three moments can be found in [Dufresne, 2001]. Since the expressions for the second and third moment are already extensive, we first test this method by use of the MQE estimates of these moments. Unfortunately, the accuracy does not improve significantly, whether the Feller condition is satisfied or not. We therefore conclude that this power series will not be of any use.

In order to increase calibration accuracy, we may try to match higher moments as well. After adding a third moment to the method, we do not notice any improvement of the calibration accuracy. As the numerical experiments suggest, the inaccuracy of the estimation of E_{ij} is again the main problem. All higher moments depend on E_{ij} . We conclude that extending the moment matching technique to multiple moments may help provided that we find a good approximation for E_{ij} .

6.6 Conclusion and outlook

In this chapter, we designed a market price implied calibration method for the multi-asset Heston model. We assumed the correlation structure of the MQE method in Chapter 4.

For the one-dimensional Heston calibration, considerations regarding efficiency and pitfalls of this method are given. This calibration can be cumbersome since the parameter space of the minimization problem is five-dimensional. The key to an efficient and accurate calibration is the insight in the derivative that is to be priced. Knowing the sensitivity of the derivative to each parameter, one can improve calibration efficiency by anticipating the discretized parameter space on the more important parameter. Furthermore, one can perform calibration in more than one step, thus solving subsequent subproblems of lower dimensionalities.

For the stock-stock correlation calibration method, market quotes of basket options for different strikes and maturities are required as input. We have tested the method in case of two stocks, measuring its performance by the error between the imposed stock-stock correlation and the calibrated stock-stock correlation. We considered 3 cases with increasing volatility of volatility parameter, using either a Black-Scholes or a Heston calibration method. Whereas the Black-Scholes calibration is much faster than the Heston calibration, both methods attain the same level of accuracy in these three cases. High accuracy is attained in case *Feller Satisfied*, medium accuracy is attained in *Feller Almost Violated*, and in case *Feller Violated*, the method accuracy is unacceptably low. We investigated this inaccuracy and concluded that the key to the solution will be an estimate for E_{ij} in (6.7) which is accurate for all practical parameter sets.

In many practical situations, the number of underlyings of a basket option will be more than 2. In case of many stock-stock calibrations, the calibration method should solve the optimization problem in an efficient way. In our method, this means that we have to find

$$\min_{\Sigma \in \mathcal{C}_D} \sum_{T \in \mathcal{T}} |\hat{M}_2^{Ss}(T) - M_2^F(T)|. \quad (6.9)$$

Here $\mathcal{C}_D$ represents the set of all correlation matrices of dimension D and $\mathcal{T}$ is the set of maturities of all used market price quotes. Using the first estimate of Section 6.3.1 can write the approximation for $\mathbb{E}[S_s^2(T)]$ as

$$\hat{M}_2^{Ss}(T) = e^{2 \int_0^T r(t) dt} \sum_{i=1}^D \sum_{j=1}^D S_i(0) S_j(0) f_{ij}, \quad \text{with} \quad f_{ij} := \exp \left\{ \rho_{ij} \sqrt{\theta_i \theta_j} T \right\}.$$

Since $\hat{M}_2^{Ss}(T)$ is linear in f_{ij} , we can use a linear optimization method to find the solution of (6.9). The simplex method is an efficient and well-known method to solve linear optimization problems (see [Dantzig et al., 2003]). The optimal solution set $\{f_{ij}, i, j = 1, \dots, D\}$ will immediately yield the optimal correlation parameter set $\{\rho_{ij}, i, j = 1, \dots, D\}$, by definition of f_{ij} . Note that for all choices of our proposed estimates for E_{ij} in Section 6.3.1, the optimization problem will be linear.

Another possible problem will be that in practice, market price quotes of the used basket options are often not liquid. This means that the basket option is traded for a number of maturities and strikes which is insufficient for accurate calibration use. We say that the derivative is not liquid. One may work around this problem by calibrating the stock-stock correlation in multiple steps:

1. We choose a subset of stocks, so that the corresponding basket option is liquid. We calibrate the corresponding stock-stock correlations by use of our proposed method.
2. Some correlations are still to be calibrated. We try to find another subset of stocks so that the corresponding basket option is liquid. We calibrate the corresponding stock-stock correlations.

3. We repeat step 2 either until all correlations are calibrated, or until there are no more liquid basket options available for the calibration of the remaining correlations. Any remaining correlation parameter can be calibrated using historical data, for example by use of the method in [Dimitroff et al., 2009].
4. Since we calibrate the stock-stock correlation with different sets of data, our correlation matrix may be positive indefinite. In that case, we try to find a 'close by' positive definite matrix, for example by use of the procedure described in [Jäckel, 2002].

Chapter 7

Conclusions

This chapter summarizes the conclusions and recommendations for further investigation of this thesis.

The goal of this thesis was to construct a multi-asset Heston Monte Carlo method. The method has to be efficient with regard to accuracy and CPU time. Additionally, we investigated the design of a market implied calibration technique for the multi-dimensional model. The following issues were encountered.

Negative variance

Conventional Taylor based schemes are not suitable to discretize Heston dynamics. Especially when the Feller condition is violated, the variance path of these scheme can become negative. This either breaks down the scheme or induces bias in derivative price estimates. Therefore, we investigated several alternative schemes for the single-asset Heston model. Eventually, the QE scheme became the scheme to use for our multi-dimensional model. Its intuitive and close approximation of the exact distribution of the next step variance makes the scheme fast and accurate, regardless of the fulfilment of the Feller condition.

Correlation imposition

We assume general stock-stock correlations and corresponding stock-variance correlations. We take the variance-variance correlations and the cross-term stock-variance correlations to be zero. Whereas the sampling of the variance processes occurs independently, the issue of 'leaking correlation' exists in a 'naive' stock propagation scheme. We avoid this leaking correlation by converting the conventional scheme so that all input random variables can be sampled independently. The resulting Monte Carlo method is the Multi-dimensional Quadratic Exponential (MQE) method.

A numerical correlation test of the MQE method is performed. It appears precisely when the Feller condition is violated, the log-stock return correlation estimate is substantially different from the imposed stock-stock correlation on the Brownian motions. This phenomenon is not observed under Black-Scholes dynamics and is theoretically different from the concept of leaking correlation. Though it is in accordance with the Heston dynamics when Feller is violated, this difference is not intuitive and may be confusing. For example, the problem arises that, when comparing the multi-asset price fairly with, say, a multi-asset Black-Scholes price, it is hard to make sure that the log-stock return correlations mutually coincide.

Efficiency and accuracy

In order to test the validation performance of the MQE method, we had to find a reference method. When the Feller condition is satisfied, the MQE prices coincide with the multi-dimensional Euler Full Truncation (FT) prices. Since the Euler FT method may inherit significant bias in case the Feller condition is violated, we had to find another reference for our MQE method. We found a semi-analytic pricing method for a double-asset option, based on a change of numéraire. We had to choose proportional variance processes and one of the two stock-variance correlation had to be taken zero. The MQE prices coincide with these prices, while the Euler FT method exhibits significant bias for long maturity and high volatility of volatility. Further, the MQE method is superior to the Euler FT method in accuracy, when fixing either the step size or the CPU time. The CPU time of the MQE method grows linear in the number of

assets of the model.

Market implied calibration method

In literature, we only found a multi-asset Heston calibration method based on historical data. Therefore, we investigated the construction of an market implied calibration method for the MQE method. The calibration method consists of D separate one-dimensional Heston model calibrations, whereafter the $D(D-1)/2$ stock-stock correlations are calibrated using a moment matching technique. For the one-dimensional Heston calibration, considerations regarding efficiency and pitfalls of this method are given. The key point to an efficient calibration is the anticipation of the optimization to the more important parameters.

The stock-stock correlation calibration method uses basket option prices for different stocks and maturities as data. It appears that the performance of the method decreases in the volatilities of volatility. The main cause is an approximation error which increases exponentially in these Heston parameters.

Further investigation

We expect that the MQE scheme can be extended by dropping the assumption of zero cross-term stock-variance correlations. The same can be said about the variance-variance correlations, though then an approximation error will be made. To correlate the different variance processes, one has to use an inexact method to correlate the concerning random variables. The resulting bias should be investigated.

Another consideration of these extensions is the increase in complexity of the required calibration method. One should either design a new method to calibrate all multi-dimensional correlations, or calibrate the added non-zero correlations and solve the problem of possible positive indefiniteness of the resulting correlation matrix.

The performance of the current stock-stock correlation method should be improved. In particular, an accurate estimate of E_{ij} in (6.7) should be conceived. Further, since in practice one may have to calibrate high-dimensional models, an efficient optimization algorithm should be applied. For example, choosing the estimate for E_{ij} appropriately, the optimization problem becomes a linear one, for which efficient solution methods are known (see Section 6.6).

Bibliography

- [Ahrens & Dieter, 1982] Ahrens, J. H. & Dieter, U. (1982). Computer generation of poisson deviates from modified normal distributions. *ACM Transactions on Mathematical Software*, 8(2), 163–179.
- [Andersen, 2007] Andersen, L. B. (2007). Efficient Simulation of the Heston Stochastic Volatility Model. *SSRN eLibrary*.
- [Bakshi et al., 1997] Bakshi, G., Cao, C., & Chen, Z. (1997). Empirical performance of alternative option pricing models. *Journal of Finance*, 52(5), 2003–2049.
- [Bendel & Mickey, 1978] Bendel, R. & Mickey, M. (1978). Population correlation matrices for sampling experiments. *Communications in Statistics-Simulation and Computation*, 7(2), 163–182.
- [Bin, 2007] Bin, C. (2007). Calibration of the Heston Model with Application in Derivative Pricing and Hedging.
- [BIS, 2009] BIS (2009). Amounts outstanding of over-the-counter derivatives. Bank for International Settlements, <http://www.bis.org/statistics/derstats.htm>.
- [Björk, 2009] Björk, T. (2009). Chapter 24 (change of numeraire). In *Arbitrage theory in continuous time*. Oxford University Press.
- [Black & Scholes, 1973] Black, F. & Scholes, M. S. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–54.
- [Broadie & Kaya, 2006] Broadie, M. & Kaya, O. (2006). Exact simulation of stochastic volatility and other affine jump diffusion processes. *Oper. Res.*, 54(2), 217–231.
- [Carr & Madan, 1999] Carr, P. & Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2(4), 61–73.
- [Chen, 2001] Chen, H. (2001). Initialization for NORTA: Generation of random vectors with specified marginals and correlations. *INFORMS Journal on Computing*, 13(4), 312.
- [Cont, 2001] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1, 223–236.
- [Cox et al., 1985] Cox, J. C., Ingersoll, Jr., J. E., & Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, Vol. 53, No. 2, pp. 385–407.
- [Dantzig et al., 2003] Dantzig, G., Orden, A., & Wolfe, P. (2003). The generalized simplex method for minimizing a linear form under linear inequality restraints. *The basic George B. Dantzig*. Stanford University Press, California, USA.
- [Dimitroff et al., 2009] Dimitroff, G., Lorenz, S., & Szimayer, A. (2009). A Parsimonious Multi-Asset Heston Model: Calibration and Derivative Pricing. *SSRN eLibrary*.
- [Duffie et al., 2000] Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), 1343–1376.

- [Dufresne, 2001] Dufresne, D. (2001). The integrated square-root process. *Research Paper*, 90.
- [Dupire, 1994] Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1), 18–20.
- [Fang & Oosterlee, 2008] Fang, F. & Oosterlee, C. W. (2008). A novel pricing method for European Options based on Fourier-cosine Series Expansions. *SIAM J. Sci. Comput.*, 31(2), 826–848.
- [Feller, 1951] Feller, W. (1951). Two singular diffusion problems. *The Annals of Mathematics*, 54, 173–182.
- [Ghosh & Henderson, 2003] Ghosh, S. & Henderson, S. (2003). Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 13(3), 276–294.
- [Glasserman & Kim, 2008] Glasserman, P. & Kim, K. (2008). Gamma Expansion of the Heston Stochastic Volatility Model. *SSRN eLibrary*.
- [Grzelak & Oosterlee, 2010] Grzelak, L. A. & Oosterlee, C. W. (2010). On the Heston Model with Stochastic Interest Rates. *Delft University of Technology Technical Report No. 09-05*.
- [Heston, 1993] Heston, S. L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, Vol. 6, No. 2 (1993), pp. 327–343.
- [Jäckel, 2002] Jäckel, P. (2002). *Monte Carlo methods in finance*. John Wiley & Sons.
- [Kirkpatrick, 1984] Kirkpatrick, S. (1984). Optimization by simulated annealing: Quantitative studies. *Journal of Statistical Physics*, 34(5), 975–986.
- [Levy, 1992] Levy, E. (1992). Pricing European average rate currency options. *Journal of International Money and Finance*, 11(5), 474–491.
- [Lord et al., 2008] Lord, R., Koekkoek, R., & van Dijk, D. J. (2008). A Comparison of Biased Simulation Schemes for Stochastic Volatility Models. *SSRN eLibrary*.
- [Mikhailov & Nögel, 2003] Mikhailov, S. & Nögel, U. (2003). Hestons stochastic volatility model implementation, calibration and some extensions. *Wilmott*, 4, 74–79.
- [Moodley, 2005] Moodley, N. (2005). The Heston Model: A Practical Approach.
- [Shreve, 2004] Shreve, S. (2004). *Stochastic Calculus for Finance: Continuous-Time Models*. Springer Verlag.
- [Turnbull & Wakeman, 1991] Turnbull, S. & Wakeman, L. (1991). A quick algorithm for pricing European average options. *Journal of Financial and Quantitative Analysis*, 26(03), 377–389.
- [van Haastrecht & Pelsser, 2008] van Haastrecht, A. & Pelsser, A. A. (2008). Efficient, Almost Exact Simulation of the Heston Stochastic Volatility Model. *SSRN eLibrary*.
- [Zhu, 2008] Zhu, J. (2008). A Simple and Exact Simulation Approach to Heston Model. *SSRN eLibrary*.

Appendix A

Cholesky decomposition

In this appendix, we give details of a Cholesky decomposition.

Theorem 2. Cholesky decomposition

Suppose that an $n \times n$ matrix Σ is real, symmetric and positive definite. Then we can write Σ as $LL^\top$, with L an $n \times n$ lower triangular matrix. This decomposition is called a Cholesky decomposition of Σ .

Before proving this theorem, we will show one of the various algorithms to find matrix L , the *Cholesky algorithm*:

1. Set $i = 1$ and $\Sigma_i = \Sigma$.
2. For $i = 1, 2, \dots$, the square matrix Σ_i of dimension $(n + 1 - i)$ has the following form:

$$\Sigma_i = \begin{pmatrix} I_{i-1} & 0 & 0 \\ 0 & a_{i,i} & \mathbf{b}_i^\top \\ 0 & \mathbf{b}_i & B_i \end{pmatrix},$$

where I_i is the identity matrix of dimension i , $a_{i,i} \in \mathbb{R}$, $\mathbf{b}_i \in \mathbb{R}^{n-i}$ a column vector and B_i a square matrix of dimension $(n - i)$. If we now define the lower triangular matrix L_i by

$$L_i = \begin{pmatrix} I_{i-1} & 0 & 0 \\ 0 & \sqrt{a_{i,i}} & 0 \\ 0 & \frac{1}{\sqrt{a_{i,i}}} \mathbf{b}_i & I_{n-i} \end{pmatrix},$$

then we can write Σ_i as $\Sigma_i = L_i \Sigma_{i+1} L_i^\top$, where

$$\Sigma_{i+1} = \begin{pmatrix} I_{i-1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & B_i - \frac{1}{a_{i,i}} \mathbf{b}_i \mathbf{b}_i^\top \end{pmatrix}.$$

3. We repeat the previous step for $i = 1, \dots, n$, obtaining $\Sigma_{n+1} = I$ in this way. Therefore we found

$$\Sigma = L_1 L_2 \dots L_n L_n^\top \dots L_2^\top L_1^\top.$$

As a result, by defining

$$L = L_1 L_2 \dots L_n,$$

(which is lower triangular since all matrices on the right-hand side are lower triangular) we have found the Cholesky decomposition of $\Sigma = LL^\top$.

Proof. We already showed that matrix L we obtained from the above algorithm is lower triangular. To show that the decomposition is always possible for symmetric, positive definite matrices, we only have to show that all computations in the algorithm are well-defined. Therefore basically, we have to show that $a_{i,i} > 0$ for all i . To do so, we define the unit vector $\mathbf{e}_i \in \mathbb{R}^n$ with elements $(\mathbf{e}_i)_j = \delta_{ij}$ for $i, j = 1, \dots, n$ and δ_{ij} the Kronecker delta. Then, since Σ is positive definite, we deduce

$$a_{1,1} = \mathbf{e}_1^\top \Sigma \mathbf{e}_1 > 0.$$

This shows that the computations are well-defined for $i = 1$. Then we claim that if Σ_i is positive definite, then the matrix $B_i - \frac{1}{a_{i,i}} \mathbf{b}_i \mathbf{b}_i^\top$ is positive definite too. To show this, we will use the property that since Σ_i is symmetric positive definite, every principal submatrix of Σ_i is symmetric and positive definite too. Now take any column vector $\mathbf{v} \in \mathbb{R}^{n-i} \setminus \{\mathbf{0}\}$ and $\mathbf{w} = -(1/a_{i,i}) \mathbf{b}_i^\top \mathbf{v}$. Then

$$\mathbf{v}^\top \left(B_i - \frac{1}{a_{i,i}} \mathbf{b}_i \mathbf{b}_i^\top \right) \mathbf{v} = \begin{pmatrix} \mathbf{w} & \mathbf{v}^\top \end{pmatrix} \begin{pmatrix} a_{i,i} & \mathbf{b}_i^\top \\ \mathbf{b}_i & B_i \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{v} \end{pmatrix} > 0.$$

The equality above follows from simple linear algebra and the inequality holds since

$$\begin{pmatrix} a_{i,i} & \mathbf{b}_i^\top \\ \mathbf{b}_i & B_i \end{pmatrix}$$

is a principal submatrix of Σ_i and is thus positive definite. Since $B_i - \frac{1}{a_{i,i}} \mathbf{b}_i \mathbf{b}_i^\top$ is positive definite, $a_{i+1,i+1}$, which is the upper left element of this matrix, is strictly positive too. By mathematical induction, we have proven that $a_{i,i} > 0$ for all i . We conclude that the Cholesky decomposition exists. $\square$

Example 1. Consider the following correlation matrix

$$\Sigma = \begin{pmatrix} 1 & \rho_{V_1 V_2} & \rho_1 & 0 \\ \rho_{V_1 V_2} & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & \rho_{12} \\ 0 & \rho_2 & \rho_{12} & 1 \end{pmatrix}.$$

Then the Cholesky algorithm yields the representation $\Sigma = LL^\top$, with

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \rho_{V_1 V_2} & \sqrt{1 - \rho_{V_1 V_2}^2} & 0 & 0 \\ \rho_1 & -\frac{\rho_{V_1 V_2} \rho_1}{\sqrt{1 - \rho_{V_1 V_2}^2}} & \sqrt{\frac{\rho_{V_1 V_2}^2 + \rho_1^2 - 1}{\rho_{V_1 V_2}^2 - 1}} & 0 \\ 0 & \frac{\rho_2}{\sqrt{1 - \rho_{V_1 V_2}^2}} & \frac{\rho_{12} - \rho_{V_1 V_2}^2 \rho_{12} + \rho_{V_1 V_2} \rho_1 \rho_2}{\sqrt{(\rho_{V_1 V_2}^2 - 1)(1 - \rho_{V_1 V_2}^2 - \rho_1^2)}} & L_{44} \end{pmatrix},$$

with

$$L_{44} = \sqrt{1 + \frac{\rho_2^2}{\rho_{V_1 V_2}^2 - 1} + \frac{(\rho_{12} - \rho_{V_1 V_2}^2 \rho_{12} + \rho_{V_1 V_2} \rho_1 \rho_2)^2}{(\rho_{V_1 V_2}^2 - 1)(1 - \rho_{V_1 V_2}^2 - \rho_1^2)}}. \quad (\text{A.1})$$

Appendix B

Derivation of the second moment of $S_s(T)$

In this appendix, we will derive the second moment of $S_s(T)$, with $S_s(t)$ as defined in (6.5). Suppose that all D stocks follow the Heston dynamics as given in (2.1) and (2.2). Then, by linearity of expectations, one has

$$\mathbb{E}[S_s^2(T)] = \sum_{i=1}^D \sum_{j=1}^D \mathbb{E}[S_i(T)S_j(T)] \quad (\text{B.1})$$

In order to find an expression for the summand, we first define

$$X(t) = \ln S_i(t) + \ln S_j(t).$$

Note that

$$S_i(T)S_j(T) = e^{X(T)} = e^{X_0 + \int_0^T dX(t)} = S_i(0)S_j(0)e^{\int_0^T dX(t)}. \quad (\text{B.2})$$

Using two-dimensional Itô's Lemma, we find

$$dX(t) = 2r(t)dt - \frac{1}{2}(V_i(t) + V_j(t))dt + \sqrt{V_i(t)}dW_{S_i}(t) + \sqrt{V_j(t)}dW_{S_j}(t). \quad (\text{B.3})$$

Since $dW_{S_i}(t)$ and $dW_{S_j}(t)$ have correlation ρ_{ij} , we can write

$$dW_{S_j}(t) \stackrel{d}{=} \rho_{ij}dW_{S_i}(t) + \sqrt{1 - \rho_{ij}^2}dW(t), \quad (\text{B.4})$$

with $W(t)$ a Brownian motion independent of everything else. Substitution of (B.4) in (B.3) and integrating the result, yields

$$\begin{aligned} \int_0^T dX(t) &= 2 \int_0^T r(t)dt - \frac{1}{2} \int_0^T (V_i(t) + V_j(t))dt + \int_0^T \left(\sqrt{V_i(t)} + \rho_{ij}\sqrt{V_j(t)} \right) dW_{S_i}(t) \\ &\quad + \sqrt{1 - \rho_{ij}^2} \int_0^T \sqrt{V_j(t)} dW(t). \end{aligned}$$

Note that given $\mathcal{V}_{ij} = \{V_i(t), V_j(t), 0 \leq t \leq T\}$, all terms on the right-hand side are independent. Therefore,

$$\begin{aligned} \mathbb{E}\left[e^{\int_0^T dX(t)} \middle| \mathcal{V}_{ij}\right] &= \exp\left\{2 \int_0^T r(t)dt - \frac{1}{2} \int_0^T (V_i(t) + V_j(t))dt\right\} \\ &\quad \times \mathbb{E}\left[\exp\left\{\int_0^T \left(\sqrt{V_i(t)} + \rho_{ij}\sqrt{V_j(t)}\right) dW_{S_i}(t)\right\} \middle| \mathcal{V}_{ij}\right] \\ &\quad \times \mathbb{E}\left[\exp\left\{\sqrt{1 - \rho_{ij}^2} \int_0^T \sqrt{V_j(t)} dW(t)\right\} \middle| \mathcal{V}_{ij}\right]. \end{aligned} \quad (\text{B.5})$$

Both exponents in the two expectations on the right-hand side are Itô integrals. Both Itô integrals are normally distributed given $\mathcal{V}_{ij}$. We can use the following rule:

Result 6. Suppose Y is normally distributed. Then

$$\mathbb{E}[e^Y] = \exp\{\mathbb{V}ar(Y)/2\}.$$

Applying this rule yields:

$$\begin{aligned}\mathbb{E}\left[\exp\left\{\int_0^T (\sqrt{V_i(t)} + \rho_{ij}\sqrt{V_j(t)})dW_{S_i}(t)\right\}\middle|\mathcal{V}_{ij}\right] &= \exp\left\{\frac{1}{2}\int_0^T V_i(t) + \rho_{ij}^2 V_j(t) + 2\rho_{ij}\sqrt{V_i(t)}\sqrt{V_j(t)}dt\right\} \\ \mathbb{E}\left[\exp\left\{\sqrt{1-\rho_{ij}^2}\int_0^T \sqrt{V_j(t)}dW(t)\right\}\middle|\mathcal{V}_{ij}\right] &= \exp\left\{\frac{1-\rho_{ij}^2}{2}\int_0^T V_j(t)dt\right\}.\end{aligned}$$

After substitution of these two expressions in (B.5), we obtain

$$\mathbb{E}\left[e^{\int_0^T dX(t)}\middle|\mathcal{V}_{ij}\right] = \exp\left\{2\int_0^T r(t)dt + \int_0^T \rho_{ij}\sqrt{V_i(t)}\sqrt{V_j(t)}dt\right\}.$$

Since $r(t)$ is deterministic,

$$\mathbb{E}\left[e^{\int_0^T dX(t)}\right] = e^{2\int_0^T r(t)dt}\mathbb{E}\left[\exp\left\{\int_0^T \rho_{ij}\sqrt{V_i(t)}\sqrt{V_j(t)}dt\right\}\right]. \quad (\text{B.6})$$

Finally, by substituting of (B.2) in (B.1), and then substituting (B.6) in the result, we arrive at

$$\mathbb{E}[S_S^2(T)] = e^{2\int_0^T r(t)dt} \sum_{i=1}^D \sum_{j=1}^D S_i(0)S_j(0)\mathbb{E}\left[\exp\left\{\int_0^T \sqrt{V_i(t)}\rho_{ij}\sqrt{V_j(t)}dt\right\}\right].$$