Vanna-Volga and Karasinski Risk Correction Methods

Ming Tao

Merton College

University of Oxford

A Thesis Submitted for The Degree of

Doctor of Philosophy

in Mathematics (Mathematical finance)

Trinity 2009

Dedication

I would like to dedicate this thesis to my mother who is always with us in our hearts, my father, and my sister.

Acknowledgements

I would like to give my heart-felt thanks to my supervisors Prof Mike Giles and Prof Sam Howison, and my corporate supervisor Dr Piotr Karasinski for the continuous supervision and unfaltering support on my research. Prof Giles and Prof Howison have poured their time and attention in a generous way to help me learn my area of research and write my research in clarity. Based on my own career goals and interests, they have encouraged me to explore the real world mathematical finance topics through an Oxford-HSBC project jointly set up with Dr Piotr Karasinski, then Managing Director of Quant Research of HSBC, who helped secure for me a three-year full D.Phil research studentship from HSBC and steered me to the main research topic of this thesis, the Vanna-Volga and Karasinski methods. It has also helped me in a great way in studying the practical topics of my thesis that Dr Karasinski made continuous effort to make available the resources of his team to benefit my research. Prof Giles, Prof Howison, and Dr Karasinski have played a big part in my thesis, for which I shall forever remain grateful.

I would like to give a big thank-you to Professor William Shaw and Dr Christoph Reisinger, who have provided invaluable suggestions and remarks that have proved to be right on point and have made my thesis improve immensely.

I would also like to thank Dr Zhongmin Qian, Mr Marouane Benchekroun, Dr Charlie Chamberlain, Prof Terry Lyons, Dr Ben Hambly, Dr Daniel Miao, Dr Alexey Polishchuk, Dr Peter Carr, Mr Vincent Craignou for helpful discussions. I am indebted to Dr Daniel Miao and Dr Klaus E. Schmitz Abe for supplying their Heston Model Monte Carlo codes for me to study and compare.

This research is funded by HSBC, and an internship at which has helped me gain valuable understanding about the practical side of my research.

Statement of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or accepted for the award of any other degree or diploma at University of Oxford or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at Oxford or elsewhere, is explicitly acknowledged in the thesis.

Other than the references cited in the thesis, I would like to acknowledge the following contributions made by others.

In the numerical methods chapters 2 and 3, Prof Mike Giles made me aware of the relevant literature and kindly supplied some of his own codes for me to learn and compare. The mesh-adaptive upwinding method for non-uniform grid was extended from the mesh-adaptive upwinding scheme for uniform grid first used by Prof Giles in finance. Prof William Shaw first reminded me the benefit of the Douglas Scheme, which has been included as part of these chapters. Based on Prof Shaw's suggestion, the numerical operators used in these chapters follow Mitchell and Griffiths [73] and Shaw [90]. The numerical results presented in Section (3.2.1) serve to present Giles & Carter results [37] in summary.

In the chapter 4 on Multi-Heston methods, Dr Piotr Karasinski first made me aware of the work of Kainth & Saravanamuttu [54]. Prof William Shaw introduced me to the highly useful "Fundamental Transform" idea by Lewis [65]. Prof Shaw suggested that the Multi-Heston model can be in fact expressed by a single variance factor. Dr. Karasinski and Prof Giles made me aware of the numerical method by Kahl & Jäckel [53].

In the chapters 5, 6, and 7 on the VV and Karasinski methods, Prof William Shaw made me aware of the Lewis expansion [65] of the Heston price into the Black-Scholes prices. Dr Alexey Polishchuk from Bloomberg commented to me that this Lewis expansion can be further expanded w.r.t ρ into the Polishchuk and Carr expansion [82]. Dr Christoph Reisinger suggested to use pseudo-inverse to better explain some of the thoughts in the chapter on the Karasinski method. Dr Reisinger suggested the drop of the "non-arbitrage

argument" and the inclusion of the discussion on whether the $(d\sigma)^2$ and $d\sigma dS$ terms should be included into the dt term in computing the VV weights. Dr Reisinger also suggested to include a hedging experiment to test the quality of the VV weights as hedging strategies in a simulated market. Prof Giles suggested to consider the complexity and difficulty of knowing for sure the market parameters of $(d\sigma)^2$ and $d\sigma dS$ terms and form an argument to support the formula to compute the VV weights. Prof Giles and Dr Karasinski suggested to look at BS-BS and Heston-BS as the assumed market-model relation for testing. Prof William Shaw and Prof Sam Howison also suggested to use Heston-BS as the assumed market-model relation for testing. Prof Shaw also suggested the CEV model for testing. Prof Howison suggested to start explaining the VV method using just one hedging instrument first. Dr Reisinger suggested to make an assumption of the Black-Scholes implied vol being a diffusion process in the discussion of the VV methods. Prof Howison further suggested to make a wider assumption of the Black-Scholes implied vol being a function of diffusion processes in the discussion of the VV method. Prof Howison suggested that the FLQ calibration method presented in the chapter on the Karasinski method can be solved linearly in approximation. Dr Karasinski pointed out that using the balance of sensitivities w.r.t. model parameters should be considered a special case of the general framework of the Karasinski method; there may be other ways to compute the price correction although this special case using the balance of sensitivities w.r.t. model parameters offers computational benefits.

Abstract

The Vanna-Volga (VV) method has been in wide use as one of the major tools for several years among foreign exchange (FX) trading desks. Despite its popularity, the properties of the VV method are not well studied and understood. This thesis attempts to understand better why and when the VV method makes sense, and how to use it better. Often under practical circumstances the state of calibration can be described as being frequent but imperfect. To take advantage of this level of calibration, we studied the properties and benefits of the Karasinski method, and extended this method to a few useful applications. We have found that the Karasinski method, if used with a reasonably calibrated model, can provide significant performance improvement over the VV method. The VV and Karasinski chapters contain most of the original research in this thesis; there are a wealth of discoveries made in these chapters. Novel methods and applications related to the VV and Karasinski methods are proposed, and some of which can be readily applied to the practical trading environment.

To make the VV and Karasinski methods work well in practice, the numerical issues for computing the price and Greeks have been carefully addressed with finite difference schemes that are second-order convergent and fast to compute. As an example of easy-to-compute but difficult-to-calibrate model candidates for the Karasinski method, the Multi-Heston model has been discussed too. A sound computational preparation enables the VV and in particular Karasinski methods to enjoy high viability as being fast, efficient and practical.

This thesis is tailored to the purpose of making a detailed study on these useful methods whose great potential has not been adequately understood and fully realised.

Contents

CHAPTER 1

Overview

This thesis is mainly concerned with fast calculation of option prices and Greeks using the Vanna-Volga (VV) and Karasinski methods, mainly for the FX market, although the results can also be applied to other markets. In the VV method, Vanna means the second-order sensitivity ($\frac{\partial^2 V^{BS}}{\partial S \partial \sigma}$) of the Black-Scholes option price V^{BS} with respect to both the underlying price S and the Black-Scholes implied volatility σ and Volga means the second-order sensitivity ($\frac{\partial^2 V^{BS}}{\partial \sigma^2}$) of the Black-Scholes option price with respect to the Black-Scholes implied volatility σ. The VV method is already very popular among FX trading desks. A characteristic of the FX market is that this market is volatile with shifting smile characteristics while traders demand quick answers with little appetite for complexity. In such a market, to use a well-calibrated but complex model to price can be expensive. The primary motivation driving the use of the VV method is that no calibration is needed in pricing options with this method. A different but broader set of problems in a similar vein initially raised by Karasinski [55] is how, if the model is not the flat-vol Black-Scholes but some general model having been imperfectly calibrated, the risk correction based method behaves to improve the accuracy.

Chapters 5, 6, and 7 contribute to the study of the VV and Karasinski methods. Despite the popularity of the VV method in FX trading, this method has not been well explained and justified. There has been very little research covering the VV and Karasinski methods. Chapter 5 first gives an introduction on the FX conventions, and proceed to explain what the VV method is by first constructing a hedging portfolio based on one liquid vanilla and then extending this portfolio to include three liquid vanillas. A discussion is dedicated to the assumptions on the spot and the Black-Scholes implied volatility made by the Castagna and Mercurio's analysis. The derivation of the equivalence between the VV price and the RRBF-VV price is made, to which the Traders' Rule price is shown just an approximation. The VV price is tested under different market assumptions such as the Heston, SABR, CEV, Merton Jump models. The performance of the VV method applied to exotics is shown to be influenced by the unbalanced Vanna terms. Chapter 6 first raises the issue

of the P factor, which modifies the VV formula. The unmodified VV method has price inconsistency when the barrier is hit. The P factor is shown to make the P-VV price correct in two extreme ends of the no-touch probability $P_{NoTouch}$. We then proceed to analyse whether the P factor makes sense. Two issues are identified to be related to the need for a P factor. The so-called V-VV method is introduced with the implications discussed. Numerical tests are done to compare various methods. A detailed discussion on the benefits of methods are given afterwards. Chapter 7 discusses a generalisation of the VV method - the Karasinski method. We first give discussion on how the Karasinski method compares to the VV method conceptually. In the following discussion, we show that the Karasinski method assumes a "parameter-error" view and the VV method assumes a "hedging-costs" view. These two views are related to the model-market relationship with which these methods are put into use. The Karasinski method is extended to include any number of liquid vanillas and any number of weights-computing equations. The resulting pseudo-inverse issue is discussed. A method of computing Greeks with the Karasinski method is proposed. A fast linear calibration method is proposed to work with the Karasinski method. Numerical analysis then follows to support the arguments made in the chapter.

Chapters 2, 3, and 4 contribute to the study of the computational methods which may benefit the VV and Karasinski methods. The VV and Karasinski approaches demand fast numerical implementation. In order to compute the model prices and Greeks in a fast and efficient fashion for the risk correction methods, we look into the relationship between discretisation, time marching and numerical convergence of the finite-difference method, which holds a speed advantage over the Monte-Carlo method in low dimensions. To compute the VV and Karasinski prices, some model Greeks will need to be computed along with the model price. For models that lack analytical solutions, second order numerical convergence ensures that the prices and Greeks can be obtained quickly. This thesis analyses the effects of various schemes on the order of convergence and the shifting pattern of convergence in different scenarios. Non-uniform grids and adaptive upwinding are studied with an emphasis on the links between them. The Multi-Heston model is studied to highlight the type of difficult-to-calibrate but fast-to-compute models that can take advantage of the Karasinski method.

Chapter 8 summarises the findings presented in the previous chapters, and discusses potential directions for future research.

CHAPTER 2

Discretisation, Non-Uniform Grids and Adaptive Upwinding

2.1. Finite Difference Methods for PDEs

2.1.1. Pros and Cons of Monte-Carlo vs. Finite-Difference Methods (FDM). The Monte-Carlo method is used to solve stochastic differential equations (SDEs) while the finite-difference method is used to obtain solutions for partial differential equations (PDEs). Since expectations of SDEs and solutions of PDEs are closely related by results such as the Feynman-Kac theorem, both methods are employed to address the needs of similar problems under many circumstances. However, other than one's taste or personal liking, there are both advantages and disadvantages associated with either method.

The Monte-Carlo method is conceptually straightforward in the basic mathematics it requires. Many software packages are available from both commercial sources or in-house development. To develop new programs is relatively easy because a new programme may be adapted from existing programs without much effort. Other benefits of Monte-Carlo solution include easy modelling of correlations and path dependency. For very high dimensional problems (dimension $d > 4$), the Monte-Carlo method is generally better than FDM because it avoids the property of FDM that the number of mesh points in FDM increases exponentially with d.

In comparison with FDM, despite the above strengths, the Monte-Carlo technique has disadvantages in the following areas. First, for dimensions up to 4, the Monte-Carlo method is slow. The error ε between the Monte-Carlo estimate and the true model value is roughly of the order of the inverse square root of the number N of simulations ($\varepsilon \sim 1/\sqrt{N}$)). Therefore an accuracy improvement by factor of 10 must be achieved by performing 100 times more simulations.

Secondly, the FDM method is superior to Monte-Carlo method in free boundary problems. In dealing with American option, one needs to optimise early exercise as t varies. To perform such a task requires all values calculated up to expiry, not just today's value. However, the Monte-Carlo method is usually used for calculation at a single point in S,

t-space. FDM when used in backward induction naturally computes the history of values as the computation moves backward in time, thus having an advantage.[1] Also the finite difference method can be better to get the Greeks.

2.1.2. Parabolic Equation and Discretisation. Many models in financial mathematics lead to linear parabolic partial differential equations,[2] which take the general form in the one-dimensional case as follows:

$$\frac{\partial V}{\partial \tau} = a\left(x,\tau\right)\frac{\partial^2 V}{\partial x^2} + b\left(x,\tau\right)\frac{\partial V}{\partial x} + c\left(x,\tau\right)V \tag{2.1.1}$$

where $\tau = T - t$ is the time to maturity, T is the maturity and $a\left(x,t\right)$ is strictly positive. In order to solve the equation, an initial condition and boundary conditions are required. From (2.1.1), we see finite difference discretisation is needed for the first and second derivatives V_τ, V_x, and V_{xx}.

Three basic forms of discretisation are forward, backward and central differences.

$$\text{Forward difference: } \frac{\partial V}{\partial x} \approx \frac{\delta V}{\delta x} = \frac{V\left(x+\delta x\right) - V(x)}{\delta x}$$

$$\text{Backward difference: } \frac{\partial V}{\partial x} \approx \frac{\delta V}{\delta x} = \frac{V\left(x\right) - V(x-\delta x)}{\delta x}$$

$$\text{Central difference: } \frac{\partial V}{\partial x} \approx \frac{\delta V}{\delta x} = \frac{V\left(x+\delta x\right) - V(x-\delta x)}{2\delta x}$$

$$\text{Central difference: } \frac{\partial^2 V}{\partial x^2} \approx \frac{\delta^2 V}{\delta x^2} \triangleq \frac{V\left(x+\delta x\right) - 2V(x) + V(x-\delta x)}{\left(\delta x\right)^2}$$

A Taylor series expansion of the above differences will show that the forward and backward differences are first order accurate and central differences are second order accurate.

Multi-dimensional models will further incur cross-derivative terms. The parabolic PDE in two dimensions can be written as

$$\frac{\partial V}{\partial \tau} = a_1\frac{\partial^2 V}{\partial x_1^2} + a_2\frac{\partial^2 V}{\partial x_2^2} + d_{12}\frac{\partial^2 V}{\partial x_1 \partial x_2} + b_1\frac{\partial V}{\partial x_1} + b_2\frac{\partial V}{\partial x_2} + cV$$

[1] See [38], [100] for related topics in this section.

[2] See [75], [94] for topics related to finite difference methods in this and the following sections.

The following central difference scheme for a cross derivative provides an example of a discretisation that is second order accurate in both x_1 and x_2:

$$\frac{\partial^2 V}{\partial x_1 \partial x_2} \approx \left(V\left(x_1+\delta x_1, x_2+\delta x_2\right) - V\left(x_1+\delta x_1, x_2-\delta x_2\right) - V\left(x_1-\delta x_1, x_2+\delta x_2\right) + V\left(x_1-\delta x_1, x_2-\delta x_2\right)\right)/4\delta x_1 \delta x_2$$

A successful numerical scheme needs to satisfy two fundamental conditions: consistency and stability. A numerical scheme is said to be consistent if the finite difference representation converges to the PDE as the space and time steps tend to zero. A numerical scheme is considered stable if perturbations to the initial conditions produce a uniformly bounded perturbation to the solution. A stable scheme only guarantees that the numerical solution will not blow up as time goes to infinity. A scheme is said to converge if the difference between the numerical solution and the exact solution at a fixed point in the domain of interest tends to zero uniformly as the space and time discretisations tend to zero.

Consistency, stability and convergence are related through the Lax Equivalence Theorem [62, 75], which states that given a consistent difference approximation to a well-posed linear initial value problem, stability is necessary and sufficient for convergence. The Lax Equivalence Theorem illustrates the importance of focusing on and understanding stability.

2.1.3. Definition of Operators. Before the introduction of numerical schemes, it is convenient to define a few operators. In defining the operators, we mainly follow Mitchell and Griffiths [73] and Shaw [90].

The differentiation operators are defined as

$$\begin{aligned} LV &= \frac{\partial V}{\partial \tau} \\ DV &= \frac{\partial V}{\partial x} \end{aligned}$$

and the difference operators are defined as

$$\begin{aligned} \delta_x V_j^n &= V_{j+1/2}^n - V_{j-1/2}^n \\ \delta_{2x} V_j^n &= V_{j+1}^n - V_{j-1}^n, \\ \delta_x^2 V_j^n &= V_{j+1}^n - 2V_j^n + V_{j-1}^n. \end{aligned} \tag{2.1.2}$$

where n is the temporal index and j the spatial index.

Assuming the Taylor expansion of V with respect to time or space is possible, the Taylor expansion of $V_j^{n+1} = V(x_j, \tau_n + \delta\tau)$ gives

$$\begin{aligned} V_j^{n+1} &= \left(1 + \delta\tau L + \frac{1}{2}(\delta\tau)^2 L^2 + ...\right) V(x_j, \tau_n) \\ &= \exp(\delta\tau L) V(x_j, \tau_n) \end{aligned} \tag{2.1.3}$$

and the Taylor expansion of $V_{j+1}^n = V(x_j + \delta x, \tau_n)$ gives

$$\begin{aligned} V_{j+1}^n &= \left(1 + \delta x D + \frac{1}{2}(\delta x)^2 D^2 + ...\right) V(x_j, \tau_n) \\ &= \exp(\delta x D) V(x_j, \tau_n) \end{aligned} \tag{2.1.4}$$

where $\delta\tau$ is the time interval and δx the grid interval.

An exact formula [44, 90, 73] linking D and δ_x is

$$D = \frac{2}{\delta x}\sinh^{-1}\left(\frac{\delta_x}{2}\right) = \frac{1}{\delta x}\left(\delta_x - \frac{1^2}{2^2 3!}\delta_x^3 + \frac{1^2 3!}{4^4 5!}\delta_x^5 ...\right) \tag{2.1.5}$$

Similarly, we can write the formula linking D and δ_{2x} as

$$D = \frac{1}{\delta x}\sinh^{-1}(\delta_{2x}) = \frac{1}{2\delta x}\left(\delta_{2x} - \frac{1^2}{2^2 3!}\delta_{2x}^3 + \frac{1^2 3!}{4^4 5!}\delta_{2x}^5 ...\right) \tag{2.1.6}$$

The formula for D^2 is given by Shaw [90]:

$$D^2 = \frac{1}{(\delta x)^2}\left(\delta_x^2 - \frac{1}{12}\delta_x^4 + \frac{1}{90}\delta_x^6 - \frac{1}{560}\delta_x^8 + \frac{1}{3150}\delta_x^{10} + O(\delta_x)^{11} ...\right) \tag{2.1.7}$$

The Taylor expansion of $\delta_{2x} V_j^n$ and $\delta_x^2 V_j^n$ are given by

$$\begin{aligned} \delta_{2x} V_j^n &= \left(2(\delta x) D + \frac{1}{3}(\delta x)^3 D^3 + O\left((\delta x)^5 D^5\right)\right) V(x_j, \tau_n) \\ \delta_x^2 V_j^n &= \left((\delta x)^2 D^2 + \frac{1}{12}(\delta x)^4 D^4 + O\left((\delta x)^6 D^6\right)\right) V(x_j, \tau_n) \end{aligned} \tag{2.1.8}$$

2.1.4. Explicit Euler Scheme. In the explicit method, (2.1.1) is discretised using only the second-order terms δ_{2x} in (2.1.6) and δ_x^2 in (2.1.7):

$$\begin{aligned} V_j^{n+1} - V_j^n \approx a(x_j, \tau_n)\frac{\delta\tau}{(\delta x)^2}\delta_x^2 V_j^n + b(x_j, \tau_n)\frac{\delta\tau}{2\delta x}\delta_{2x} V_j^n \\ + \delta\tau\, c(x_j, \tau_n) V_j^n \end{aligned} \tag{2.1.9}$$

Plugging (2.1.8) into (2.1.9) gives

$$\frac{V_j^{n+1} - V_j^n}{\delta\tau} - LV\left(x_j,\, \tau_n\right) \approx$$
$$\left[-\frac{1}{2}\left(\delta\tau\right) L^2 + \left(\delta x\right)^2 \left(\frac{a(x_j,\tau_n)}{12} D^4 + \frac{b(x_j,\tau_n)}{6} D^3\right) ...\right] V\left(x_j,\, \tau_n\right),$$

hence the truncation error of this scheme is $O(\delta\tau, (\delta x)^2)$.

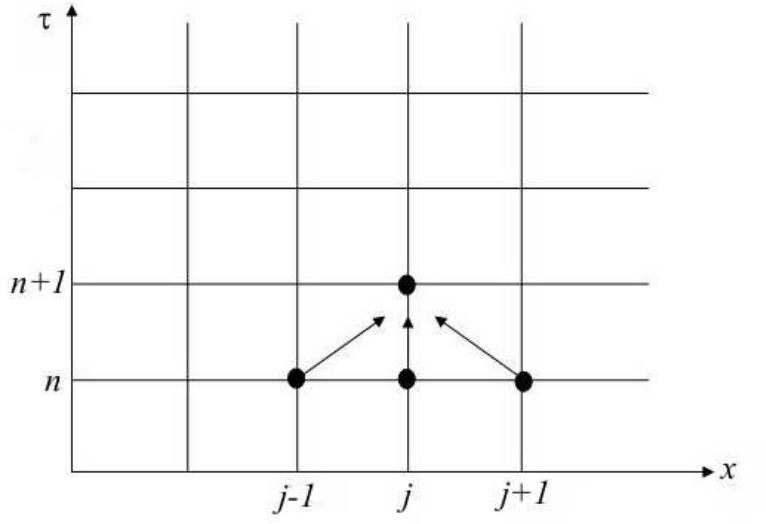

FIGURE 2.1.1. Explicit Euler Discretisation

The explicit method calculates the values at time $n+1$ using only the values at time n. Therefore computation can be done easily by progressing forwards in time to maturity backwards in calendar time. This method can become unstable if $\delta\tau / \left(\delta x\right)^2$ is too large. When the size of δx needs to be reduced to increase the accuracy of solution, we need to sharply reduce $\delta\tau$ as well. Therefore this scheme could be costly in computation time.

Assuming $a = \max\left(a\left(x_j, \tau_n\right)\right)$, the stability requirement for the explicit scheme is $0 < \delta\tau \leqslant \left(\delta x\right)^2 / \left(2a\right)$.

2.1.5. Backward Euler Scheme. A Backward Euler scheme replaces the forward time difference with a backward time difference, and the space differences remain the same.

$$V_j^{n+1} - V_j^n \approx a(x_j, \tau_n + \delta\tau)\frac{\delta\tau}{\left(\delta x\right)^2}\delta_x^2 V_j^{n+1} +$$
$$b(x_j, \tau_n + \delta\tau)\frac{\delta\tau}{2\delta x}\delta_{2x} V_j^{n+1} + \delta\tau\, c(x_j, \tau_n + \delta\tau) V_j^{n+1} \quad (2.1.10)$$

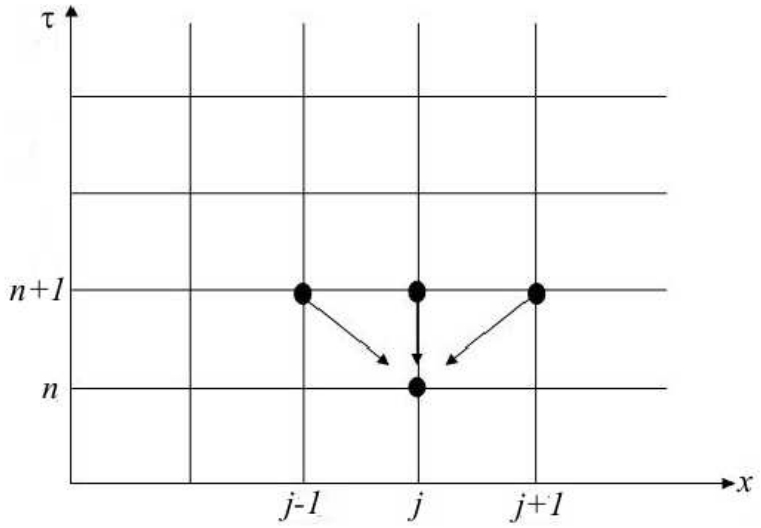

FIGURE 2.1.2. Implicit Euler Discretisation

The Backward Euler method is unconditionally stable; ie no restriction is needed in $\delta\tau$. Obviously it is an improvement over the explicit method in this regard. So the Backward Euler method can take much larger time steps than the explicit method[75]. Nonetheless, this method is still a first order method with respect to time because the truncation error of this scheme is $O(\delta\tau, (\delta x)^2)$.

This scheme is unconditionally stable ($\delta\tau > 0$).

2.1.6. Crank-Nicolson Scheme. The Crank-Nicolson method is a mix of the explicit and the Backward Euler methods. It is a special case of the θ- method. Let $0 < \theta < 1$; we can use θ and $1 - \theta$ to obtain a weighted average of (2.1.9) and (2.1.10) of the form

$$\begin{aligned} V_j^{n+1} - V_j^n = \theta &\left[a\left(x_j, \tau_n + \delta\tau\right) \frac{\delta\tau}{(\delta x)^2} \delta_x^2 V_j^{n+1} + b\left(x_j, \tau_n + \delta\tau\right) \frac{\delta\tau}{2\delta x} \delta_{2x} V_j^{n+1} \right. \\ &\left. + \delta\tau\, c\left(x_j, \tau_n + \delta\tau\right) V_j^{n+1} \right] \\ + (1-\theta) &\left[a\left(x_j, \tau_n\right) \frac{\delta\tau}{(\delta x)^2} \delta_x^2 V_j^n + b\left(x_j, \tau_n\right) \frac{\delta\tau}{2\delta x} \delta_{2x} V_j^n + \delta\tau\, c\left(x_j, \tau_n\right) V_j^n \right] \end{aligned} \quad (2.1.11)$$

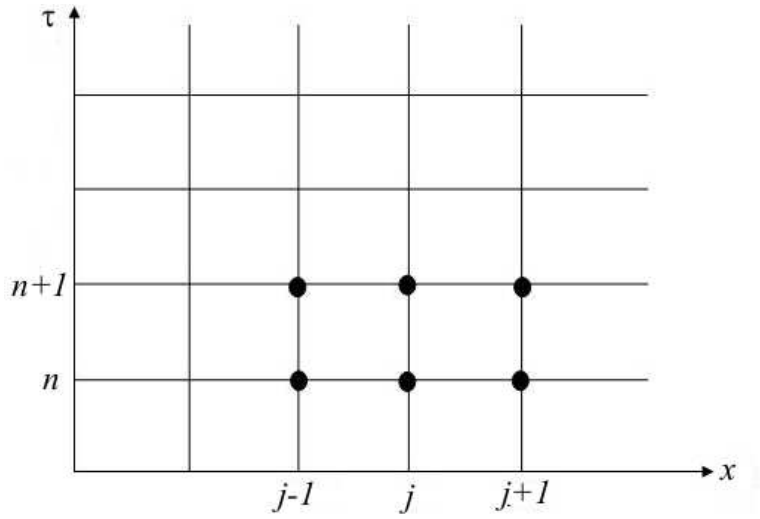

FIGURE 2.1.3. Crank-Nicolson Discretisation

It is clear that (2.1.11) is an explicit method when $\theta = 0$, and a Backward Euler method when $\theta = 1$. Both methods incur a discretisation error of the order of $O(\delta\tau, (\delta x)^2)$. When $\theta = 1/2$, (2.1.11) becomes the Crank-Nicolson scheme. Truncation error analysis shows that the Crank-Nicolson scheme is second order with respect to both time and space $(O((\delta\tau)^2, (\delta x)^2))$. Obviously having second-order convergence represents a significant improvement over both explicit and Backward-Euler methods, which are only first-order methods. Furthermore, just like the Backward-Euler method, the Crank-Nicolson method is unconditionally stable.[3] Because of these two nice properties, the Crank-Nicolson scheme is a suitable and powerful tool for many problems in financial mathematics. This scheme is unconditionally stable ($\delta\tau > 0$).

However, the Crank-Nicolson scheme has an undesirable effect of oscillation of its solution around the strike when the payoff function is not smooth. To continue to enjoy the benefit of the Crank-Nicolson scheme, we can combine some other methods with this scheme in order to achieve satisfactory results.

2.1.7. Discretisation of the Black-Scholes Equation. For the purpose of convenience, when the Black-Scholes equation is used in our research, we often transform the co-ordinates t and S to τ and x. Assuming S ranges from 0 to S_{max}, we let $S = S_{max}x$, $\tau = T - t$. Through this transformation, mathematical backward-marching becomes

[3]This is actually a simplified statement, which will be discussed later in the thesis.

forward-marching in computer code, and the payoff function becomes the initial condition. The domain of the equation now becomes $[0,1]$. If the domain is discretised into J grid points, we will have $\delta x = 1/J$. The Black-Scholes equation

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS\frac{\partial V}{\partial S} - rV = 0 \tag{2.1.12}$$

now takes the form

$$\frac{\partial V}{\partial \tau} = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 V}{\partial x^2} + rx\frac{\partial V}{\partial x} - rV \tag{2.1.13}$$

After discretising (2.1.13) using the explicit method, we have

$$V^{n+1} - V^n = \frac{1}{2}\delta\tau\,\sigma^2 j^2 \delta_x^2 V_j^n + \frac{1}{2} r \delta\tau j \delta_{2x} V_j^n - \delta\tau\, r\, V_j^n \tag{2.1.14}$$

If we discretise (2.1.13) using the Backward Euler method, we will have

$$V^{n+1} - V^n = \frac{1}{2}\delta\tau\,\sigma^2 j^2 \delta_x^2 V_j^{n+1} + \frac{1}{2} r\,\delta\tau\, j\,\delta_{2x} V_j^{n+1} - \delta\tau\, r\, V_j^{n+1} \tag{2.1.15}$$

The average of (2.1.14) and (2.1.15) will yield the Crank-Nicolson form

$$\begin{aligned} V^{n+1} - V^n = {} & \tfrac{1}{2}[\tfrac{1}{2}\delta\tau\,\sigma^2 j^2 \delta_x^2 V_j^{n+1} + \tfrac{1}{2}\delta\tau\, r\, j\,\delta_{2x} V_j^{n+1} - \delta\tau\, r\, V_j^{n+1}] \\ & + \tfrac{1}{2}[\tfrac{1}{2}\delta\tau\,\sigma^2 j^2 \delta_x^2 V_j^{n} + \tfrac{1}{2}\delta\tau\, r\, j\,\delta_{2x} V_j^{n} - \delta\tau\, r\, V_j^{n}]. \end{aligned} \tag{2.1.16}$$

2.1.8. Douglas Scheme. Shaw [90] applied the $\theta-$method as discussed in the Crank-Nicolson method section to the discretisation of the diffusion equation

$$LU = D^2 U \tag{2.1.17}$$

and gave the difference equation to the order δ_x^6

$$\begin{aligned} U_j^{n+1} - U_j^n \approx {} & \alpha\theta\delta_x^2 U_j^{n+1} + \alpha\,(1-\theta)\,\delta_x^2 U_j^n \\ & - \frac{1}{2}\theta\alpha\left(\alpha\theta + \frac{1}{6}\right)\delta_x^4 U_j^{n+1} + \alpha\theta\left(\frac{\theta^2\alpha^2}{6} + \frac{\theta\alpha}{12} + \frac{1}{90}\right)\delta_x^6 U_j^{n+1} \\ & + \left(\frac{1}{2}(1-\theta)^2\alpha^2 - \frac{1}{12}(1-\theta)\,\alpha\right)\delta_x^4 U_j^n \\ & + \left(\frac{1}{6}\alpha^2(1-\theta)^2 - \frac{1}{12}\alpha(1-\theta) + \frac{1}{90}\right)\alpha\,(1-\theta)\,\delta_x^6 U_j^n \end{aligned} \tag{2.1.18}$$

where $\alpha = \dfrac{\delta\tau}{(\delta x)^2}$.

For a general value of θ, Wilmott [100] gave the local truncation error as

$$O\left(\frac{1}{2}\delta\tau - \frac{1}{12}(\delta x)^2 - \theta\delta\tau,\, (\delta x)^4,\, (\delta\tau)^2\right).$$

Thus the truncation error will become $O\left((\delta x)^4,\, (\delta\tau)^2\right)$ when θ is chosen to be

$$\theta = \frac{1}{2} - \frac{1}{12}\frac{(\delta x)^2}{\delta\tau} \tag{2.1.19}$$

By choosing the θ by (2.1.19), this numerical scheme is called the Douglas scheme [90, 100]. Mitchell & Griffiths [73] suggest that the optimal $\dfrac{\delta\tau}{(\delta x)^2}$ is

$$\frac{\delta\tau}{(\delta x)^2} = \frac{1}{\sqrt{20}} \tag{2.1.20}$$

because this choice of $\dfrac{\delta\tau}{(\delta x)^2}$ minimises the truncation error of the scheme.

For the Black-Scholes equation (2.1.12) with constant parameters, Wilmott et al [101] gave the following transform formula

$$V(t,\, S) = K\exp\left(-\frac{1}{2}(k-1)\,x - \frac{1}{4}(k+1)^2\,\tau\right)U(\tau,\, x)$$

where

$$\begin{aligned} LU &= D^2U,\ -\infty < x < \infty,\ \tau > 0 \\ S &= K\,e^x \\ k &= 2r/\sigma^2 \\ \tau &= \sigma^2(T-t)/2 \end{aligned}$$

After the transform, the payoff will be

$$\begin{aligned} &\max\left(e^{\frac{1}{2}(k+1)x} - e^{\frac{1}{2}(k-1)x},\ 0\right), \qquad \text{for call option} \\ &\max\left(e^{\frac{1}{2}(k-1)x} - e^{\frac{1}{2}(k+1)x},\ 0\right), \qquad \text{for put option} \end{aligned}$$

for the heat equation of U. So the Black-Scholes equation with constant parameters can be solved through the heat equation with the Douglas scheme. The Douglas scheme is unconditionally stable ($\delta\tau > 0$).

	Truncation Error	Stability Condition
Explicit Scheme:	$O\left((\delta x)^2, (\delta\tau)\right)$	$0 < \delta\tau \leqslant (\delta x)^2 / (2a)$
Implicit Scheme:	$O\left((\delta x)^2, (\delta\tau)\right)$	$\delta\tau > 0$
Crank-Nicolson:	$O\left((\delta x)^2, (\delta\tau)^2\right)$	$\delta\tau > 0$
Douglas:	$O\left((\delta x)^4, (\delta\tau)^2\right)$	$\delta\tau > 0$

TABLE 2.1. Comparison of truncation error and stability of schemes

The following table lists the truncation errors and stability[4] of these schemes that have been discussed. Because the Douglas scheme is fourth-order in space, it has a clear advantage over other schemes in terms of error convergence.

The difficulty limiting the use of Douglas scheme is that this scheme cannot be applied as a general method to an arbitrary convection-diffusion equation with non-constant parameters. For non-constant parameters, the choice of θ will usually depend on the parameters so there may be no choice of θ that can remove all the second-order terms.

2.2. Non-uniform Grids

2.2.1. Discretisation of Non-uniform Grids. For non-uniform grids, the first derivative of V can be discretised as

$$\frac{\partial V}{\partial x} \approx \frac{1}{2}\left(\frac{V_{j+1} - V_j}{\delta x_{j+1/2}} + \frac{V_j - V_{j-1}}{\delta x_{j-1/2}}\right) \tag{2.2.1}$$

or

$$\frac{\partial V}{\partial x} \approx \frac{\delta x_{j+1/2}\left(\frac{V_{j+1}-V_j}{\delta x_{j+1/2}}\right) + \delta x_{j-1/2}\left(\frac{V_j-V_{j-1}}{\delta x_{j-1/2}}\right)}{\delta x_{j+1/2} + \delta x_{j-1/2}} = \frac{V_{j+1} - V_{j-1}}{\delta x_{j+1/2} + \delta x_{j-1/2}} \tag{2.2.2}$$

where $\delta x_{j+1/2} \triangleq x_{j+1} - x_j$ and $\delta x_{j-1/2} \triangleq x_j - x_{j-1/2}$ as illustrated in Figure 2.2.1.

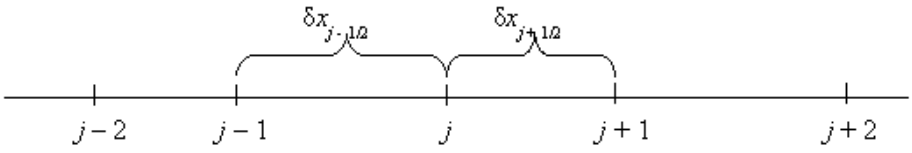

FIGURE 2.2.1. The way $\delta x_{j-1/2}$ and $\delta x_{j+1/2}$ are defined

The first discretisation expression of first derivative is basically an average of two first-order discretisations on the two sides of the point of interest. The second way of

[4]The stability issue will be revisited for the Crank-Nicolson and Douglas schemes in Chapter 3.

discretisation for the first derivative is a weighted average of two first-order discretisations. The second way of discretisation essentially computes the value of second derivative at the mid-point between the points $j-1$ and $j+1$. If the j point is significantly shifted from the mid-point between the points $j-1$ and $j+1$, being an average of the two sides, the first way of discretisation is better than the second way of discretisation. In most practical cases in this thesis, however, the two methods do not seem to differ much in accuracy.

The second derivative has only one form of discretisation

$$\frac{\partial^2 V}{\partial x^2} \approx 2\left(\frac{V_{j+1}-V_j}{\delta x_{j+1/2}} - \frac{V_j - V_{j-1}}{\delta x_{j-1/2}}\right) / \left(\delta x_{j-1/2} + \delta x_{j+1/2}\right) \tag{2.2.3}$$

When a non-uniform grid is used in practice, we usually treat a point x_j on the non-uniform grid as a function of a point y_i on a uniform grid.

PROPOSITION 1. *If the function $x(y)$ is twice differentiable on the y domain $[y_{min}, y_{max}]$, the above discretisation schemes are second order with respect to δy.*

PROOF. We can use Taylor series expansion to verify that they are second-order schemes. For the first way of discretising $\partial V/\partial x$,

$$\begin{aligned}
\frac{\partial V}{\partial x} &\approx \frac{1}{2}\left(\frac{V_{j+1}-V_j}{\delta x_{j+1/2}} + \frac{V_j - V_{j-1}}{\delta x_{j-1/2}}\right) \\
&= \frac{1}{2}\left(\frac{\delta x_{j+1/2}\, V_x|_j + \frac{1}{2}\left(\delta x_{j+1/2}\right)^2 V_{xx}|_j + ...}{\delta x_{j+1/2}} + \frac{\delta x_{j-1/2}\, V_x|_j - \frac{1}{2}\left(\delta x_{j-1/2}\right)^2 V_{xx}|_j + ...}{\delta x_{j-1/2}}\right) \\
&= \frac{1}{2}\left(V_x|_j + \frac{1}{2}\delta x_{j+1/2}\, V_{xx}|_j + ... + V_x|_j - \frac{1}{2}\delta x_{j-1/2}\, V_{xx}|_j + ...\right) \\
&= V_x|_j + \frac{1}{4}\left(\delta x_{j+1/2} - \delta x_{j-1/2}\right) V_{xx}|_j + ...
\end{aligned} \tag{2.2.4}$$

where V_x means $\frac{\partial V}{\partial x}$ precisely.

Because

$$\begin{aligned}
\delta x_{j+1/2} - \delta x_{j-1/2} &= x_{j+1} - 2x_j + x_{j-1} \\
&= x(y+\delta y) - 2x(y) + x(y-\delta y) \\
&\approx x_{yy}(\delta y)^2 .
\end{aligned} \tag{2.2.5}$$

so the first way of discretisation (2.2.1) is second order with respect to δy.

Similarly, we have

$$\begin{aligned}
\frac{\delta V}{\delta x} &= \frac{V_{j+1} - V_{j-1}}{\delta x_{j+1/2} + \delta x_{j-1/2}} \\
&= \frac{V|_j + \delta x_{j+1/2}\, V_x|_j + \frac{1}{2}\left(\delta x_{j+1/2}\right)^2 V_{xx}|_j + ... - V|_j + \delta x_{j-1/2}\, V_x|_j - \frac{1}{2}\left(\delta x_{j-1/2}\right)^2 V_{xx}|_j ...}{\delta x_{j+1/2} + \delta x_{j-1/2}} \\
&= V_x|_j + \frac{+\frac{1}{2}\left(\delta x_{j+1/2}\right)^2 V_{xx}|_j + ... - \frac{1}{2}\left(\delta x_{j-1/2}\right)^2 V_{xx}|_j ...}{\delta x_{j+1/2} + \delta x_{j-1/2}} \\
&= V_x|_j + \frac{+\frac{1}{2}\left(\delta x_{j+1/2}\right)^2 V_{xx}|_j + ... - \frac{1}{2}\left(\delta x_{j-1/2}\right)^2 V_{xx}|_j ...}{\delta x_{j+1/2} + \delta x_{j-1/2}} \\
&= V_x|_j + \frac{1}{2}\left(\delta x_{j+1/2} - \delta x_{j-1/2}\right) V_{xx}|_j + ... \qquad (2.2.6)
\end{aligned}$$

and so this scheme is also second order.

For second derivatives, we have

$$\begin{aligned}
\frac{\delta^2 V}{\delta x^2} &= 2\left(\frac{V_{j+1} - V_j}{\delta x_{j+1/2}} - \frac{V_j - V_{j-1}}{\delta x_{j-1/2}}\right) / \left(\delta x_{j-1/2} + \delta x_{j+1/2}\right) \\
&= \frac{2}{\delta x_{j-1/2} + \delta x_{j+1/2}} \left(\frac{\delta x_{j+1/2}\, V_x|_j + \frac{1}{2}\left(\delta x_{j+1/2}\right)^2 V_{xx}|_j + \frac{1}{8}\left(\delta x_{j+1/2}\right)^3 V_{xxx}|_j ...}{\delta x_{j+1/2}}\right. \\
&\left. - \frac{\delta x_{j-1/2}\, V_x|_j - \frac{1}{2}\left(\delta x_{j-1/2}\right)^2 V_{xx}|_j + \frac{1}{8}\left(\delta x_{j-1/2}\right)^3 V_{xxx}|_j ...}{\delta x_{j-1/2}}\right) \\
&= \frac{2}{\delta x_{j-1/2} + \delta x_{j+1/2}} \left(\frac{\delta x_{j-1/2} + \delta x_{j+1/2}}{2} V_{xx}|_j + \frac{\left(\delta x_{j+1/2}\right)^2 - \left(\delta x_{j-1/2}\right)^2}{8} V_{xxx}|_j + ...\right) \\
&= V_{xx}|_j + \frac{1}{4}\left(\delta x_{j+1/2} - \delta x_{j-1/2}\right) V_{xxx}|_j ... \qquad (2.2.7)
\end{aligned}$$

Therefore, as above, this scheme is second order. □

Because x (y) has a role in determining the order and magnitude of the truncation error, it is important to choose a good functional form for it. Figure (2.2.2) gives an example of the effects on error by applying non-uniform grids to the Black-Scholes equation (2.1.12). The parameters for the equation are $r = 0.02$, $\sigma = 0.2$, $T = 1$, $K = 1$ and $S = 1$. The error in the plot is the maximum absolute error between the numerical price and the Black-Scholes formula price for a put option over the range of $[0, 5]$. While the overall convergence rate is similar, the error of the non-uniform grid $x = y + y^3$ is smaller than the uniform grid in this case is because this non-uniform grid is coarser towards the end of

the range $[0, 5]$ and finer close to the strike 1. For different problems, the choices of best performing non-uniform grids will vary.

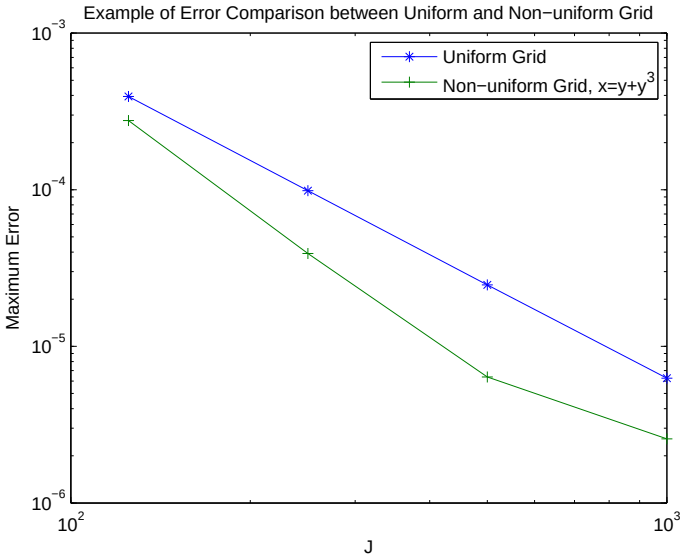

FIGURE 2.2.2. Non-uniform grid applied to put option with the Black-Scholes model

2.2.2. Coordinate Transform and Non-uniform Grids. Coordinate transform is often used to place denser grids on critical areas such as around the strike, and coarser grids on areas where the solution is flat.

Given the following general convection diffusion equation,

$$\frac{\partial V}{\partial t} + a(t, y)\frac{\partial V}{\partial y} + c(t, y)\frac{\partial^2 V}{\partial y^2} = f(t, y)V, \tag{2.2.8}$$

we define the coordinate transform as

$$x = x(y), \tag{2.2.9}$$

which implies

$$\begin{aligned} \frac{\partial V}{\partial y} &= \frac{\partial V}{\partial x}\frac{\partial x}{\partial y} \\ \frac{\partial^2 V}{\partial y^2} &= \frac{\partial^2 V}{\partial x^2}\left(\frac{\partial x}{\partial y}\right)^2 + \frac{\partial V}{\partial x}\frac{\partial^2 x}{\partial y^2} \end{aligned} \tag{2.2.10}$$

Therefore, (2.2.8) can be written as

$$\frac{\partial V}{\partial t} + a(t, y(x))\, x_y \frac{\partial V}{\partial x} + c(t, y(x))\left((x_y)^2\frac{\partial^2 V}{\partial x^2} + x_{yy}\frac{\partial V}{\partial x}\right) = f(t, y(x))V \tag{2.2.11}$$

hence

$$\begin{aligned}\frac{\partial V}{\partial t}+\left(a\left(t,y\left(x\right)\right)x_y+c\left(t,y\left(x\right)\right)x_{yy}\right)\frac{\partial V}{\partial x}\\ +c\left(t,y\left(x\right)\right)\left(x_y\right)^2\frac{\partial^2 V}{\partial x^2}=f\left(t,y\left(x\right)\right)V\end{aligned} \tag{2.2.12}$$

For a two dimensional convection diffusion equation

$$\begin{aligned}\frac{\partial V}{\partial t}+a\left(t,y_1,y_2\right)\frac{\partial V}{\partial y_1}+b\left(t,y_1,y_2\right)\frac{\partial V}{\partial y_2}+c\left(t,y_1,y_2\right)\frac{\partial^2 V}{\partial y_1^2}\\ +d\left(t,y_1,y_2\right)\frac{\partial^2 V}{\partial y_2^2}+e\left(t,y_1,y_2\right)\frac{\partial^2 V}{\partial y_1\partial y_2}=f\left(t,y_1,y_2\right)V\end{aligned}\ , \tag{2.2.13}$$

we define the coordinate transform as

$$\begin{aligned}x_1 &= x_1\left(y_1,y_2\right)\\ x_2 &= x_2\left(y_1,y_2\right)\end{aligned} \tag{2.2.14}$$

We can obtain

$$\begin{aligned}\frac{\partial V}{\partial y_1} &= \frac{\partial V}{\partial x_1}\frac{\partial x_1}{\partial y_1}+\frac{\partial V}{\partial x_2}\frac{\partial x_2}{\partial y_1}\\ \frac{\partial V}{\partial y_2} &= \frac{\partial V}{\partial x_1}\frac{\partial x_1}{\partial y_2}+\frac{\partial V}{\partial x_2}\frac{\partial x_2}{\partial y_2}\\ \frac{\partial^2 V}{\partial y_1^2} &= \left(\frac{\partial^2 V}{\partial x_1^2}\frac{\partial x_1}{\partial y_1}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_2}{\partial y_1}\right)\left(\frac{\partial x_1}{\partial y_1}\right)+\frac{\partial V}{\partial x_1}\frac{\partial^2 x_1}{\partial y_1^2}\\ &\quad+\left(\frac{\partial^2 V}{\partial x_2^2}\frac{\partial x_2}{\partial y_1}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_1}{\partial y_1}\right)\left(\frac{\partial x_2}{\partial y_1}\right)+\frac{\partial V}{\partial x_2}\frac{\partial^2 x_2}{\partial y_1^2}\\ \frac{\partial^2 V}{\partial y_2^2} &= \left(\frac{\partial^2 V}{\partial x_1^2}\frac{\partial x_1}{\partial y_2}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_2}{\partial y_2}\right)\left(\frac{\partial x_1}{\partial y_2}\right)+\frac{\partial V}{\partial x_1}\frac{\partial^2 x_1}{\partial y_2^2}\\ &\quad+\left(\frac{\partial^2 V}{\partial x_2^2}\frac{\partial x_2}{\partial y_2}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_1}{\partial y_2}\right)\left(\frac{\partial x_2}{\partial y_2}\right)+\frac{\partial V}{\partial x_2}\frac{\partial^2 x_2}{\partial y_2^2}\\ \frac{\partial^2 V}{\partial y_1\partial y_2} &= \left(\frac{\partial^2 V}{\partial x_1^2}\frac{\partial x_1}{\partial y_2}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_2}{\partial y_2}\right)\left(\frac{\partial x_1}{\partial y_1}\right)+\frac{\partial V}{\partial x_1}\frac{\partial^2 x_1}{\partial y_1\partial y_2}\\ &\quad+\left(\frac{\partial^2 V}{\partial x_2^2}\frac{\partial x_2}{\partial y_2}+\frac{\partial^2 V}{\partial x_1\partial x_2}\frac{\partial x_1}{\partial y_2}\right)\left(\frac{\partial x_2}{\partial y_1}\right)+\frac{\partial V}{\partial x_2}\frac{\partial^2 x_2}{\partial y_1\partial y_2}\end{aligned} \tag{2.2.15}$$

Substituting (2.2.15) to (2.2.13), we can have the coordinate transformed PDE.

A reason for doing coordinate transform for multi-dimensional problems is to eliminate cross derivative terms. In the above case, the coefficient of the $\frac{\partial^2 V}{\partial x_1\partial x_2}$ term is

$$c\frac{\partial x_2}{\partial y_1}\frac{\partial x_1}{\partial y_1}+c\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_1}+d\frac{\partial x_2}{\partial y_2}\frac{\partial x_1}{\partial y_2}+d\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_2}+e\frac{\partial x_2}{\partial y_2}\frac{\partial x_1}{\partial y_1}+e\frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1} \tag{2.2.16}$$

For some problems, it is possible to make the above expression zero by choosing the right forms for x_1 and x_2, hence eliminating the cross derivatives in the new coordinate

system. For example, if c and e are constants, we can set

$$\begin{aligned} x_1 &= y_1 \\ x_2 &= y_1 + py_2 \end{aligned} . \tag{2.2.17}$$

and substituting (2.2.17) into (2.2.16), and we find so that the coefficient of $\dfrac{\partial^2 V}{\partial x_1 \partial x_2}$ vanishes

$$p = -\frac{2c}{e}. \tag{2.2.18}$$

Substituting (2.2.18), (2.2.17), and (2.2.15) into (2.2.13), we obtain the coordinate transformed (2.2.13) without the cross derivative term,

$$\begin{aligned} &\frac{\partial V}{\partial t} + a^*(t, x_1, x_2)\frac{\partial V}{\partial x_1} + \left(a^*(t, x_1, x_2) - 2\frac{b^*(t, x_1, x_2)\, c^*}{e^*}\right)\frac{\partial V}{\partial x_2} \\ &+c^*\frac{\partial^2 V}{\partial x_1^2} + \left(c^* - 2\frac{d^*(t, x_1, x_2)\, c^*}{e^*}\right)\frac{\partial^2 V}{\partial x_2^2} = f^*(t, x_1, x_2)\, V \end{aligned} \tag{2.2.19}$$

where

$$\begin{aligned} a^*(t, x_1, x_2) &= a\left(t, x_1, \frac{x_2 - x_1}{p}\right) \\ b^*(t, x_1, x_2) &= b\left(t, x_1, \frac{x_2 - x_1}{p}\right) \\ c^* &= c \\ d^*(t, x_1, x_2) &= d\left(t, x_1, \frac{x_2 - x_1}{p}\right) \\ e^* &= e \\ f^*(t, x_1, x_2) &= f\left(t, x_1, \frac{x_2 - x_1}{p}\right) \end{aligned} \tag{2.2.20}$$

2.3. Mesh-adaptive Upwinding

2.3.1. Diagonal Dominance and Tridiagonal Solver (Thomas Algorithm). Following Mitchell & Griffiths [73], the finite difference equations for a parabolic PDE

$$-a_j V_{j-1} + b_j V_j - c_j V_{j+1} = d_j,\ 1 \leqslant j \leqslant J$$

can be written in matrix form as

$$\mathbf{AV} = \mathbf{d}$$

where

$$\mathbf{A} = \begin{pmatrix} b_1 & -c_1 & & & \\ -a_2 & b_2 & -c_2 & & \\ & ... & & & \\ & & -a_{J-1} & b_{J-1} & -c_{J-1} \\ & & & -a_J & b_J \end{pmatrix}, \mathbf{V} = \begin{pmatrix} V_1 \\ V_2 \\ ... \\ V_{J-1} \\ V_J \end{pmatrix}, \mathbf{d} = \begin{pmatrix} d_1 + a_1 V_0 \\ d_2 \\ ... \\ d_{J-1} \\ d_J + c_J V_{J+1} \end{pmatrix}$$

and V_0, V_{J+1} are known from the boundary conditions.

If $\mathbf{A}$ is diagonally dominant, $a_j > 0$, $b_j > 0$, $c_j > 0$, and $b_j \geqslant a_j + c_j$, $\mathbf{V}$ can be solved with a very efficient algorithm [73, 75] called either the tridiagonal matrix algorithm (TDMA) [21] or the Thomas algorithm [75, 21]. This algorithm is a simplified Gaussian elimination method without the error growth associated with the back substitution in the Gaussian elimination method and with less computer storage requirement [73].

Sometimes, when central differencing is used, the property of diagonal dominance of the tridiagonal matrix $\mathbf{A}$ is no longer true. In this case, upwinding can help to restore diagonal dominance for this matrix.

2.3.2. Upwinding for Uniform Grids. Rewrite the Black-Scholes equation

$$\frac{\partial V}{\partial \tau} = \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 V}{\partial x^2} + rx\frac{\partial V}{\partial x} - rV \tag{2.3.1}$$

in the form

$$\frac{\partial V}{\partial \tau} = \frac{1}{2}a_2 \frac{\partial^2 V}{\partial x^2} + a_1 \frac{\partial V}{\partial x} - rV \tag{2.3.2}$$

where $a_2 = \sigma^2 x^2$, and $a_1 = r\,x$. To obtain second order convergence in x, we use central differencing.

$$\begin{aligned} \delta_{2x} V_j^n &= V_{j+1}^n - V_{j-1}^n \\ \delta_x^2 V_j^n &= V_{j+1}^n - 2V_j^n + V_{j-1}^n \end{aligned} \tag{2.3.3}$$

However, when central differencing is used, there is a desirable condition on the size of grid interval δx for which the spatial tridiagonal matrix is diagonally dominant. For the discretised (2.3.2), this condition becomes

$$\delta x \leqslant \frac{2\left(\frac{1}{2}a_2^n\right)}{|a_1^n|} \tag{2.3.4}$$

In many practical problems, a_2^n may be significantly smaller than a_1^n. To ensure diagonal dominance, this will result in a very small δx, which translates to a large number of mesh points.

The use of upwinding can solve this issue by lifting the restriction (2.3.4). Upwinding is the differencing method with which we use the forward difference when a_1^n is positive and the backward difference when it is negative [75].

Upwinding relaxes the size requirement on δx. But it brings a downside of lowering the order of truncation error. While the central difference is $O((\delta x)^2)$, the forward difference introduces an error of $O(\delta x)$. When some mesh points need to use the forward/backward difference and the others use the central difference, the overall error will be between $O(\delta x)$ and $O((\delta x)^2)$. If only a small number of mesh points need to use the forward/backward difference, the overall error will remain $O((\delta x)^2)$ approximately.

If we implement upwinding using the definition discussed above, we will need to determine the sign of a_1^n and use many "ifs" and "thens" in numerical programmes. This could be costly to programme speed. An alternative way is normally used to achieve the same effect of upwinding. Especially when GPU (NVidia Cuda[5] and etc) computation is involved, many branching clauses will make the performance suffer significantly. The adaptive upwinding we will next discuss avoids doing the branching logic at the micro-level and instead makes upwinding happen in a batch at the beginning, hence improving computation efficiency.

There is another downside of upwinding. The purpose of implementing upwinding is to maintain diagonal dominance. Upwinding may be an overkill in some instances as so much diffusion is added to the scheme in upwinding as to reduce the order of convergence. In other words, regular upwinding may add more diffusion than necessary to keep diagonal dominance. By comparison, adaptive upwinding addresses this problem by adding just enough amount of diffusion.

2.3.3. Mesh-adaptive Upwinding for Uniform Grids. For (2.3.2), a mesh-adaptive upwinding can be implemented by replacing a_2 with

$$a_2^{adapt} = \max(a_2, |a_1|\delta x/2) \tag{2.3.5}$$

[5]http://www.nvidia.com/cuda

For easy reference, we call this type of upwinding the mesh-adaptive upwinding so as to differentiate itself with the term "adaptive" that normally implies solution-adaptive methods. This method has been used in engineering for many years (see [88, 93]).

When the convection term is larger than the diffusion term, this is essentially just enough upwinding that ensures that the discretisation is diagonally dominant and a tridiagonal solver can be used without any possibility of ill-conditioning. Oftentimes, mesh-adaptive upwinding helps maintain the second order convergence while maintaining diagonal dominance.

On the boundaries, central difference does not apply as one point is always missing. Thus forward and backward differences are employed there. The boundary condition employed is to set the diffusion term zero. This boundary condition is rather common.[94] If the range between boundaries is wide enough, it is normal to ignore the diffusion term on the boundaries. To meet these two requirements, we conveniently use $a_2^{adapt} = -a_1\delta x/2$ on the upper boundary and $a_2^{adapt} = a_1\delta x/2$ on the lower boundary.

2.3.4. Mesh-adaptive Upwinding Formula for Non-uniform Grid. The idea of mesh-adaptive upwinding is to add just enough diffusion to make the tridiagonal matrix diagonally dominant when the convection term is too big in value compared to the diffusion term. We have derived the mesh-adaptive upwinding formula for the non-uniform grid.

PROPOSITION 2. *Assume a_1 is the coefficient of the first x-derivative term and a_2 the coefficient of the second x-derivative term. For first derivative discretisation satisfying (2.2.1) or (2.2.2) and second derivative discretisation satisfying (2.2.3), adaptive upwinding can be implemented by re-adjusting a_2 to a_2^{adapt} based on a_1, the original a_2, and the grid intervals using the following formula:*

$$a_2^{adapt} = \max\left(a_2, |a_1|\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)/4\right) \tag{2.3.6}$$

if (2.2.1) is used, or

$$a_2^{adapt} = \begin{cases} \max\left(a_2, \dfrac{1}{2}|a_1|\delta x_{j-1/2}\right), \ a_1 \geqslant 0, \ \delta x_{j+1/2} < \delta x_{j-1/2} \\ \max\left(a_2, \dfrac{1}{2}|a_1|\delta x_{j+1/2}\right), \ a_1 < 0, \ \delta x_{j+1/2} \geqslant \delta x_{j-1/2} \\ \max\left(a_2, |a_1|\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)/4\right), \ a_1\left(\delta x_{j+1/2} - \delta x_{j-1/2}\right) \geqslant 0 \end{cases} \tag{2.3.7}$$

if (2.2.2) is used. For the financial PDE problems for which only the boundary condition $V_{xx}|_{boundary} = 0$ *is applied, on the boundary the value of* a_2 *needs to be re-adjusted to*

$$a_2^* = |a_1| \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)/4 \tag{2.3.8}$$

if (2.2.1) is used, or

$$a_2^* = \begin{cases} \max\left(a_2, \frac{1}{2}|a_1|\delta x_{j-1/2}\right), & a_1 \geqslant 0 \\ \max\left(a_2, \frac{1}{2}|a_1|\delta x_{j+1/2}\right), & a_1 < 0 \end{cases} \tag{2.3.9}$$

if (2.2.2) is used.

PROOF. We prove only the case for (2.2.1); the analysis for (2.2.2) is similar.

Without loss of generality, assume the option pricing PDE has the following form:

$$V_\tau = a_1 V_x + a_2 V_{xx} - rV \tag{2.3.10}$$

where $\tau = T - t$, T is the maturity, and a_2>0, $r \geqq 0$. This equation can be discretised using the Crank-Nicolson scheme as

$$V^{n+1} - V^n = \frac{\delta\tau}{2}\left(a_1 V_x^{n+1} + a_2 V_{xx}^{n+1} + rV^{n+1}\right) + \frac{\delta\tau}{2}\left(a_1 V_x^n + a_2 V_{xx}^n + rV^n\right) \tag{2.3.11}$$

We will first discretise $a_1 V_x^{n+1} + a_2 V_{xx}^{n+1}$ using (2.2.1) and (2.2.3).

$$\begin{aligned} a_1\frac{\delta V^{n+1}}{\delta x} + a_2\frac{\delta^2 V^{n+1}}{\delta x^2} \quad & \frac{a_1}{2}\left(\frac{V_{j+1}^{n+1} - V_j^{n+1}}{\delta x_{j+1/2}} + \frac{V_j^{n+1} - V_{j-1}^{n+1}}{\delta x_{j-1/2}}\right) \\ & + \frac{2a_2}{\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}\left(\frac{V_{j+1}^{n+1} - V_j^{n+1}}{\delta x_{j+1/2}} - \frac{V_j^{n+1} - V_{j-1}^{n+1}}{\delta x_{j-1/2}}\right) \\ = \quad & \frac{g_{j+1}V_{j+1}^{n+1} + g_j V_j^{n+1} + g_{j-1}V_{j-1}^{n+1}}{2\delta x_{j+1/2}\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)} \end{aligned} \tag{2.3.12}$$

where

$$\begin{aligned} g_{j+1} &= a_1\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) + 4a_2\delta x_{j-1/2} \\ g_j &= a_1\left(\left(\delta x_{j+1/2}\right)^2 - \left(\delta x_{j-1/2}\right)^2\right) - 4a_2\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ g_{j-1} &= -a_1\delta x_{j+1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) + 4a_2\delta x_{j+1/2} \end{aligned} \tag{2.3.13}$$

Note in Equation (2.3.10), we do not need to consider cross derivatives for the purpose of this proof because they are not included on the left side in the tridiagonal matrices; for

all the schemes we use in this thesis, cross derivatives if applicable are placed on the right side of finite-difference equations.

We can rearrange (2.3.11) as

$$b_{j+1}V_{j+1}^{n+1} + b_jV_j^{n+1} + b_{j-1}V_{j-1}^{n+1} = d_j^n \tag{2.3.14}$$

where

$$\begin{aligned} b_{j+1} &= -a_1\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) - 4a_2\delta x_{j-1/2} \\ b_j &= \frac{4\delta x_{j+1/2}\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}{\delta\tau} + 2\delta x_{j+1/2}\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) r \\ &\quad +4a_2\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) - a_1\left(\left(\delta x_{j+1/2}\right)^2 - \left(\delta x_{j-1/2}\right)^2\right) \\ b_{j-1} &= a_1\delta x_{j+1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) - 4a_2\delta x_{j+1/2} \end{aligned} \tag{2.3.15}$$

For convenience, we define

$$\begin{aligned} r^* = &\frac{4\delta x_{j+1/2}\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}{\delta\tau} \\ &+2\delta x_{j+1/2}\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) r \end{aligned} \tag{2.3.16}$$

which obviously is positive.

We are not concerned with the content in d_j^n because only the tridiagonal matrices represented by b_{j+1}, b_j, are b_{j-1} are relevant in our analysis. d_j^n denotes whatever the right-hand side is once the form of the left-hand side is written as in (2.3.14). So we do not need to write d_j^n in full detail here.

First, we will prove that

$$|b_j| > |b_{j+1}| + |b_{j-1}| \tag{2.3.17}$$

where

$$a_2 \geqslant \frac{|a_1|\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}{4} \tag{2.3.18}$$

If (2.3.18) holds, we obtain

$$\begin{aligned} |b_{j+1}| &= a_1\delta x_{j-1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) + 4a_2\delta x_{j-1/2} \\ |b_j| &= r^* + 4a_2\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ &\quad -a_1\left(\left(\delta x_{j+1/2}\right)^2 - \left(\delta x_{j-1/2}\right)^2\right) \\ |b_{j-1}| &= -a_1\delta x_{j+1/2}\left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) + 4a_2\delta x_{j+1/2} \end{aligned} \tag{2.3.19}$$

hence

$$\begin{aligned} |b_j| &= |b_{j+1}| + |b_{j-1}| + r^* \\ &> |b_{j+1}| + |b_{j-1}| \end{aligned} \tag{2.3.20}$$

Now we will try to establish diagonal dominance for the case of

$$a_2 < \frac{|a_1| \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}{4} \tag{2.3.21}$$

In this case, the diffusion component is not sufficient to ensure tridiagonal dominance. Therefore, some extra diffusion is added by replacing a_2 with

$$a_2^{adapt} = \frac{|a_1| \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)}{4} \tag{2.3.22}$$

Substituting $a_2 = a_2^{adapt}$ into (2.3.15), we obtain

$$\begin{aligned} |b_{j+1}| &= |\left(a_1 + |a_1|\right)| \delta x_{j-1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ |b_j| &= \Big| r^* + |a_1| \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right)^2 \\ &\quad -a_1 \left(\left(\delta x_{j+1/2}\right)^2 - \left(\delta x_{j-1/2}\right)^2\right) \Big| \\ |b_{j-1}| &= |\left(a_1 - |a_1|\right)| \delta x_{j+1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \end{aligned} \tag{2.3.23}$$

When $a_1 \geqslant 0$, (2.3.23) becomes

$$\begin{aligned} |b_{j+1}| &= 2a_1 \delta x_{j-1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ |b_j| &= r^* + 2a_1 \delta x_{j-1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ |b_{j-1}| &= 0 \end{aligned} \tag{2.3.24}$$

whereas for $a_1 < 0$, (2.3.23) becomes

$$\begin{aligned} |b_{j+1}| &= 0 \\ |b_j| &= r^* - 2a_1 \delta x_{j+1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \\ |b_{j-1}| &= -2a_1 \delta x_{j+1/2} \left(\delta x_{j+1/2} + \delta x_{j-1/2}\right) \end{aligned} \tag{2.3.25}$$

In both cases $a_1 \geqslant 0$ and $a_1 < 0$, it is obvious that $|b_j| > |b_{j+1}| + |b_{j-1}|$ holds.

We have proved that (2.3.22) adds just the right amount of extra diffusion adaptively by adjusting a_2 to ensure tridiagonal dominance when the diffusion term in the PDE is not large enough. Therefore, (2.3.6) is a suitable choice of mesh-adaptive upwinding formula.

For the problems for which natural boundary conditions are applied, the convection flow always goes out of the boundary, hence on the right boundary, $a_1 < 0$, whereas on the

left boundary, $a_1 > 0$. It is easy to verify that with (2.3.8), we have $b_{j-1} = 0$ on the left boundary and $b_{j+1} = 0$ on the right boundary. $\square$

Having just enough upwinding is ideal. After all, upwinding is a first-order scheme. Therefore, too much upwinding damages the second-order convergence of the Crank-Nicolson scheme. (2.3.6) has the advantage of being a simple form of implementation.

CHAPTER 3

Fast Numerical Schemes for Non-smooth Payoffs

3.1. Numerical Schemes of Second Order Convergence

3.1.1. Inadequate Crank-Nicolson Time Marching. To obtain price and Greeks through the PDE method is particularly useful if the numerical algorithm of our choice achieves second order convergence with respect to time and space. The success of the Vanna-Volga and Karasinski methods that will be discussed in the later chapters rely upon sound numerical algorithms capable of computing both price and Greeks in a speedy fashion. In this chapter, we will discuss schemes specifically targeted at problems with non-smooth payoffs.

The Crank-Nicolson time marching is often described as an unconditionally stable second-order numerical scheme. However, this is a simplified statement. Richtmyer & Morton [87] and Smith [92] reported that non-smooth initial conditions can induce oscillation that cannot be eliminated by refining the grids; rather the oscillation can become worse as the grids are made finer. In the case of European call or put payoff, the oscillation around the strike will be greatest when the oscillation appears. Smith [92, 90] categorised the numerical schemes using the Padé approximation labelled $P(s, t)$, where s and t are integers; When $s > t$, the associated scheme has the property of "strong stability" . Smith [92] concluded that the Crank-Nicolson scheme as being a $P(1, 1)$ scheme does not hold the property of strong stability whereas the fully implicit scheme being a $P(1, 0)$ scheme does. Richtmyer & Morton [87] and Smith [92] recommended to apply three time-steps to the Crank-Nicolson and Douglas schemes to overcome the stability issue. Chawla [18] analysed the stability problems under a generalised set of trapezoidal formulas for parabolic equations labelled GTF(α)-FDS, $\alpha \in [0, 1]$, of which the special case of $\alpha = 0$ is the Crank-Nicolson scheme. Chawla [18] suggested that the Crank-Nicolson scheme is only A-stable but not L- stable and hence it is not able to cope with the non-smooth and discontinuous initial conditions. Chawla's generalised trapezoidal formulas are second-order in space and time except for GTF($\alpha = 1/3$)-FDS, for which case Chawla's formula is

third-order in time and second-order in space. Giles and Carter [37] pointed out that the payoff of a European call/put option does have limited local $\mathcal{L}^2$ norm and so does its first derivative $\frac{\partial V}{\partial S}$. But its second derivative ($\frac{\partial^2 V}{\partial S^2}$) does not have limited local $\mathcal{L}^2$ norm. The Crank-Nicolson scheme is unconditionally stable in the $\mathcal{L}^2$ space, but only conditionally stable in the $\mathcal{L}^\infty$ space. Consequently, it is only guaranteed to converge in the $\mathcal{L}^2$ space, for $\mathcal{L}^2$ initial data. This causes the property of second-order convergence not to hold for non-smooth payoff functions.

If the payoff function is smooth everywhere, the Crank-Nicolson time marching will be quite satisfactory in terms of having the desirable effect of second order convergence. However, many financial instruments do not have smooth payoffs. Take the vanilla call or put for example. The neighbourhood around the strike is not smooth. For digital options it is even more difficult to maintain second order convergence. Therefore, the content of this chapter is more or less relevant to most of the existing financial option PDE problems, hence choosing the right schemes for fast convergence is a major not a minor issue.

3.1.2. Three-Time-Step Douglas Scheme. It was demonstrated by Shaw [90] that the Douglas scheme will result in the same oscillation as in the Crank-Nicolson scheme for the non-smooth payoff functions. Shaw [90] used three-time-step Douglas scheme to compute price and Greeks for the Black-Scholes equation and obtained excellent convergence results.

The Black-Scholes equation can be reduced to a heat equation problem as discussed in the last chapter. The three-time-step Douglas scheme for the heat equation (2.1.17) is given by

$$\left(\frac{1}{8}-\frac{\delta\tau}{(\delta x)^2}\right)\left(V_{j-1}^{n+1}+V_{j+1}^{n+1}\right)+\left(\frac{5}{4}+2\frac{\delta\tau}{(\delta x)^2}\right)V_j^{n+1}=$$
$$\frac{1}{6}\left(V_{j-1}^{n}+V_{j+1}^{n}+10V_j^{n}\right)-\frac{1}{24}\left(V_{j-1}^{n-1}+V_{j+1}^{n-1}+10V_j^{n-1}\right) \tag{3.1.1}$$

which uses the two-time-step Douglas scheme

$$\left(1-6\frac{\delta\tau}{(\delta x)^2}\right)\left(V_{j-1}^{n+1}+V_{j+1}^{n+1}\right)+\left(10+12\frac{\delta\tau}{(\delta x)^2}\right)V_j^{n+1}=$$
$$\left(1+6\frac{\delta\tau}{(\delta x)^2}\right)\left(V_{j-1}^{n}+V_{j+1}^{n}\right)+\left(10-12\frac{\delta\tau}{(\delta x)^2}\right)V_j^{n} \tag{3.1.2}$$

to kick start. The two-time-step Douglas scheme is first used once for a quarter time step, then the three-time-step Douglas scheme is used once for a quarter time step and once again for a half time step. The three-time-step Douglas scheme is used on the rest of the whole time steps.

For the problems that can be solved by the three-time-step Douglas scheme, the fourth-order error convergence is superior. However, the Douglas scheme applied to a general convection-diffusion equation is not as straightforward as the Crank-Nicolson scheme. Not all convection-diffusion equations can be converted to a standard heat equation. For some problems, the choice of θ may depend on x, t, and V, making the Douglas scheme not suitable. The same issue is with Chawla's [18] GTF($\alpha = 1/3$)-FDS scheme. The choice of $\alpha = 1/3$ was derived with the heat equation. For a general convection diffusion equation, there might not be a third-order choice of α. Even if there is one for a specific equation, this choice will likely be obtained through co-ordinate transform and new truncation error analysis for this particular equation. Overall, while the schemes like Douglas and GTF($\alpha = 1/3$)-FDS may have higher order of accuracy, they may require more tailored work for each specific problem in order to take advantage of the accuracy benefits. Nonetheless, these schemes will be highly valuable in practice to the quant desks who are specialised to certain problems that can be solved with them.

Since we are studying the numerical schemes with the purpose of using them for a general convection diffusion problem on a real trading desk, we will focus on the remedies for the second-order Crank-Nicolson scheme. These schemes can be implemented once to deal with a wide majority of the problems, hence being very easy to maintain.

3.1.3. Rannacher Start-up. To recover second-order convergence in Crank-Nicolson time-marching with discontinuous initial data, Rannacher [85] studied the $\mathcal{L}^2$ convergence issue in convection-diffusion approximations. With energy methods he proved that second-order convergence can be maintained by replacing the Crank-Nicolson approximation for the very first timestep by two half-timesteps using the Backward Euler scheme. This treatment removes the oscillation in the solution by adding some smoothness to the payoff through a couple of initial time steps with the Backward Euler method.

Giles and Carter [37] analysed this problem and proposed that replacing the first two timesteps ($R = 2$ where R denotes the number of initial time steps using the fully implicit scheme.) with four half-timesteps be regarded as optimal. Using more than four

half-timesteps will increase the overall error. Their research reveals that in the Crank-Nicolson scheme coupled with the Rannacher start-up there are two types of errors, the low wavenumber error and the high wavenumber error. The low wavenumber error is $O((\delta x)^2)$ (δx is the grid interval) and the high wavenumber error is $O((\delta x)^{2R-1})$.

With $R = 0$, which is the standard Crank Nicolson scheme without the Rannacher startup, the high wavenumber error in approximating Γ is $O((\delta x)^{-1})$. As δx becomes smaller, the high wavenumber error grows bigger. At some point, the low wavenumber error may become smaller than the high wavenumber error. When this happens, the $O((\delta x)^{-1})$ high wavenumber terms become dominant and the $O((\delta x)^2)$ convergence will not hold.

When in the Rannacher startup the first two timesteps are replaced with four half-timesteps ($R = 2$), the high wavenumber error now becomes $O((\delta x)^3)$. Therefore, in this scheme, the low wavenumber error remains dominant for all δx, and the $O((\delta x)^2)$ convergence continues to hold.

3.1.4. ADI Time-Marching. Assume the two dimensional convection-diffusion equation in the generic form

$$\frac{\partial V}{\partial \tau} = a_1 \frac{\partial^2 V}{\partial x^2} + a_2 \frac{\partial^2 V}{\partial y^2} + a_{12} \frac{\partial^2 V}{\partial x \partial y} + b_1 \frac{\partial V}{\partial x} + b_2 \frac{\partial V}{\partial y} + cV \tag{3.1.3}$$

The discretisation of (3.1.3) can be written as

$$(1 - \frac{1}{2} D_x - \frac{1}{2} D_y)\delta V^{n+1} = (1 + D_x + D_y + D_{xy})V^n \tag{3.1.4}$$

where $\delta V^{n+1} = V^{n+1} - V^n$, D_x denotes the central-difference discretisation of $\delta\tau \left(a_1 \frac{\partial^2}{\partial x^2} + b_1 \frac{\partial}{\partial x} + cV \right)$, D_y denotes the central-difference discretisation of $\delta\tau \left(a_2 \frac{\partial^2}{\partial y^2} + b_2 \frac{\partial}{\partial y} \right)$ and D_{xy} denotes the central-difference discretisation of $\delta\tau\, a_{12} \frac{\partial^2}{\partial x \partial y}$. Although cV here is included in D_x, cV can actually be included in either D_x or D_y.

If this equation is solved directly, $(J_x - 1)(J_y - 1)$ linear equations will be involved. The matrix of the system has many zeros but is not tridiagonal so it cannot be solved using the efficient Thomas algorithm. Although this equation can certainly be solved numerically, directly solving it is laborious.

The ADI time-marching was first proposed by Peaceman and Rachford [79]. This method rewrites the left-hand side operator of (3.1.4) in the functional form

$$(1-\frac{1}{2}D_x)(1-\frac{1}{2}D_y)\delta V^{n+1} = (1+D_x+D_y+D_{xy})V^n \tag{3.1.5}$$

This is a splitted version of Crank-Nicolson so that the ADI scheme can be introduced. While the classical Crank-Nicolson scheme with which the D_{xy} term is also time centred is second order, we will focus on the splitted version of the Crank-Nicolson scheme and the ADI scheme, and the easy-to-implement Craig-Sneyd scheme that will be introduced after the ADI scheme.

Compared to (3.1.4), (3.1.5) incurs an extra error term $\frac{1}{4}D_xD_y\delta V^{n+1}$, but this extra term is of the same order as the truncation error. If the left side of (3.1.4) is discretised as $(1-\frac{1}{2}D_x-\frac{1}{2}D_y)V^{n+1}$, using the ADI scheme will incur an extra error term $\frac{1}{4}D_xD_yV^{n+1}$. Generally $\frac{1}{4}D_xD_y\delta V^{n+1}$ is smaller than $\frac{1}{4}D_xD_yV^{n+1}$. So using δV^{n+1} on the left side of (3.1.4) is usually better than using V^{n+1}.

(3.1.5) can be written in an equivalent form with the introduction of an intermediate variable W

$$\begin{array}{rcl}(1-\frac{1}{2}D_x)W & = & (1+D_x+D_y+D_{xy})V^n \\ (1-\frac{1}{2}D_y)\delta V^{n+1} & = & W\end{array} \tag{3.1.6}$$

With (3.1.6), W is first solved and then δV^{n+1} is solved. The advantage of using the ADI method is that multi-factor problems now can be solved by effective tridiagonal system solver through the Thomas algorithm. There are variants (e.g. Douglas & Rachford [69], D'yakonov [30]) of the ADI scheme, which mainly differ in how the intermediate values are calculated.

Now consider the three dimensional convection-diffusion equation

$$\begin{array}{rcl}\frac{\partial V}{\partial \tau} & = & a_1\frac{\partial^2 V}{\partial x^2}+a_2\frac{\partial^2 V}{\partial y^2}+a_2\frac{\partial^2 V}{\partial z^2}+a_{12}\frac{\partial^2}{\partial x\partial y}+a_{23}\frac{\partial^2}{\partial y\partial z} \\ & & +a_{13}\frac{\partial^2}{\partial x\partial z}+b_1\frac{\partial V}{\partial x}+b_2\frac{\partial V}{\partial y}+b_2\frac{\partial V}{\partial z}+cV\end{array} \tag{3.1.7}$$

The three dimensional discretisation of (3.1.7) can be written as

$$(1-\frac{1}{2}D_x-\frac{1}{2}D_y-\frac{1}{2}D_z)\delta V^{n+1} = (1+D_x+D_y+D_z+D_{cross})V^n \tag{3.1.8}$$

where $\delta V^{n+1} = V^{n+1} - V^n$, D_x is the central-difference discretisation of $\delta\tau\left(a_1\frac{\partial^2}{\partial x^2} + b_1\frac{\partial}{\partial x}\right)$, D_y is the central-difference discretisation of $\delta\tau\left(a_2\frac{\partial^2}{\partial y^2} + b_2\frac{\partial}{\partial y}\right)$, D_z is the central-difference discretisation of $\delta\tau\left(a_3\frac{\partial^2}{\partial z^2} + b_3\frac{\partial}{\partial z}\right)$ and D_{cross} is the central-difference discretisation of $\delta\tau\left(a_{12}\frac{\partial^2}{\partial x\partial y} + a_{23}\frac{\partial^2}{\partial y\partial z} + a_{13}\frac{\partial^2}{\partial x\partial z}\right)$.

The ADI method in 3D can be written in the form

$$\begin{aligned}(1-\frac{1}{2}D_x)W_1 &= (1+D_x+D_y+D_z+D_{cross})V^n \\ (1-\frac{1}{2}D_y)W_2 &= W_1 \\ (1-\frac{1}{2}D_z)\delta V^{n+1} &= W_2\end{aligned} \tag{3.1.9}$$

3.1.5. Craig-Sneyd Correction. The Crank-Nicolson scheme is a second order scheme in both time and space $O((\delta\tau)^2) + O((\delta x)^2) + O((\delta y)^2)$ for a two-dimensional case. However, when the ADI scheme is used and the cross derivative $\frac{\partial^2}{\partial x\partial y}$ term is present, the ADI scheme becomes first order in time $O(\delta\tau) + O((\delta x)^2) + O((\delta y)^2)$. The reason for the drop of order is that because the cross derivatives cannot be split into single dimensions in the ADI method, their time differencing is one-sided. To remedy this, Craig and Sneyd [27] explored an iterative version of the ADI scheme to "time-centre" the cross derivatives.

Consider an equation in its matrix form

$$AV^{n+1} = (A+B)V^n \tag{3.1.10}$$

where

$$A = \prod_{i=1}^{N}(1-\frac{1}{2}D_{x_i}) \tag{3.1.11}$$

and

$$B = \sum_{i=1}^{N} D_{x_i} + \frac{1}{2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} D_{x_i x_j} \tag{3.1.12}$$

Define the cross derivative matrix in B

$$M = \frac{1}{2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} D_{x_i x_j} \tag{3.1.13}$$

The Craig and Sneyd method adds a second step to time-centre the cross derivatives as follows:

$$\begin{aligned}AV_{(1)}^{n+1} &= (A+B)V^n \\ AV^{n+1} &= (A+B)V^n + \tfrac{1}{2}M(V_{(1)}^{n+1} - V^n)\end{aligned} \tag{3.1.14}$$

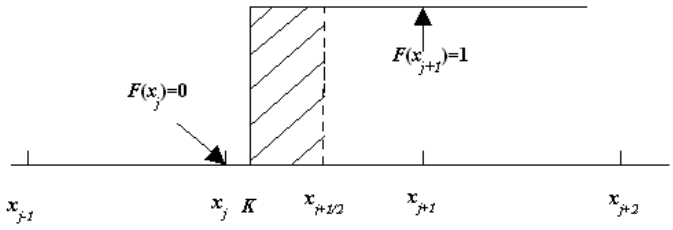

FIGURE 3.1.1. No Smoothing for discontinuous payoff

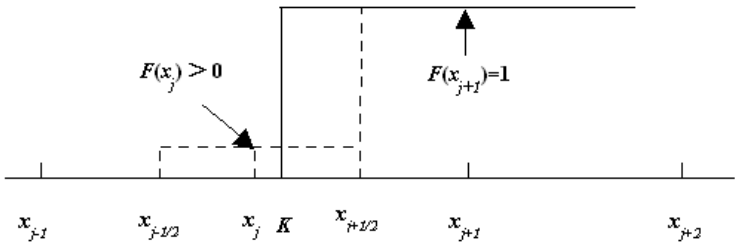

FIGURE 3.1.2. Averaging for discontinuous payoff

For 2D and 3D problems, the Craig and Sneyd method is unconditionally stable and second-order convergent in both time and space. Each step of (3.1.14) can be split by the ADI time marching method.

3.1.6. Smoothing for Discontinuous Payoffs. Figure (3.1.1) shows a digital payoff function $F(x)$ with the strike K. The strike K falls between the grid points x_j and $x_{j+1/2}$, which is the halfway point between x_j and x_{j+1}. Without any smoothing, any $K \in (x_j, x_{j+1/2}]$ will give the same discretised $F(x)$ with $F(x_j) = 0$ and $F(x_{j+1}) = 1$. Obviously an error will be introduced through this simple way of discretisation of discontinuous payoffs. Tavella and Randall [94] commented that a simple remedy that usually increases the accuracy is to shift the grids so that K falls right at $x_{j+1/2}$. It is easy to do for one dimensional problems. But grid shifting is not a general method for all problems; For example, a two-dimensional digital payoff $F(x_1 - x_2 - K)$ has no fixed discontinuous point for either dimension, hence not suitable for grid shifting.

In practice, the practioners often use the averaging method (see the definition in [94])

$$F(x_j) = \frac{1}{x_{j+1/2} - x_{j-1/2}} \int_{x_{j-1/2}}^{x_{j+1/2}} F(x_j - y)\, dy \tag{3.1.15}$$

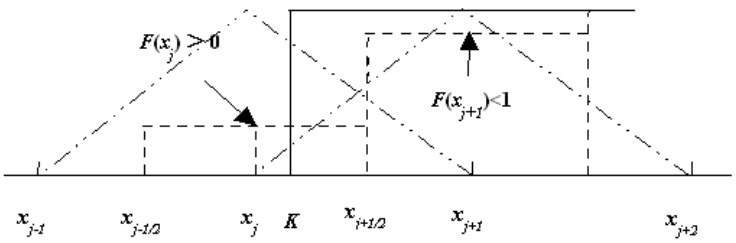

FIGURE 3.1.3. The Pooley Method for discontinuous payoff

When $F(x)$ is the digital payoff with the strike K,

$$F(x_j) = \frac{(x_{j+1/2} - K)}{(x_{j+1/2} - x_{j-1/2})}.$$

Kreiss [60] and Thomée & Wahlbin [95] have proven the validity of the averaging method. Heston and Zhou [43] showed that it increases the rate of convergence. It was demonstrated by Pooley et al [83] that the averaging method results in second-order convergence.

Pooley et al [83] proposed a method of smoothing the payoffs, which interpolates the payoff function using nodal approximations and consequently achieves second-order convergence for the integration error. The Pooley method is a special case of the projection method (see Rannacher [85]) with the lumped mass matrix approximation (see [45, 35]).

Let N_i be the standard Lagrange basis function, or the "hat" function. On the nodes,

$$\begin{aligned} N_i &= 1 \quad \text{at node } i \\ &= 0 \quad \text{at all other nodes} \\ \textstyle\sum_k N_k &= 1 \quad \text{everywhere in the solution domain} \end{aligned} \tag{3.1.16}$$

$N_i(x)$ between node i and node $i+1$ (or node $i-1$) is simply defined by a line linking these two nodes.

F_k is defined as

$$F_k = (\delta x)^{-1} \int N_k(x) F(x) dx \tag{3.1.17}$$

where $F(x)$ is the payoff function.

Pooley et al [83] proved that the resulting integration error is second order,

$$\sum_k p(x_k, T) \delta x F_k - \int p(x, T) F(x) dx = O((\delta x)^2) \tag{3.1.18}$$

where p is the state price probability density function and δx is the grid interval.

When a strike is perfectly located at the mid point of a grid interval, for piecewise constant payoffs, we can choose not to use any smoothing treatment because in this case no smoothing is the same as the averaging method. Since the averaging method is already second-order, the cost of applying a smoothing technique can be saved.

The Pooley method appears to be more complex than the averaging method. But there seems to be an error advantage over the averaging method in some occasions as Pooley et al [83] demonstrated in a two-factor numerical test. It may be due to the fact that the Pooley method in general makes a discontinuous payoff smoother than the averaging method. For example, if $K = x_{j+1/2}$, the grid points x_j and x_{j+1} in Figure (3.1.2) will have the values 0 and 1 respectively, and the grid points x_j and x_{j+1} in Figure (3.1.3) will have the values 1/8 and 7/8 respectively.

One might argue that processing a payoff across all grid points with the Pooley method could bring additional computational cost. Actually, the Pooley treatment is often needed for a very limited set of grid points around the strike, and separation of dimensions allows quick computation for many multi-dimensional payoffs.

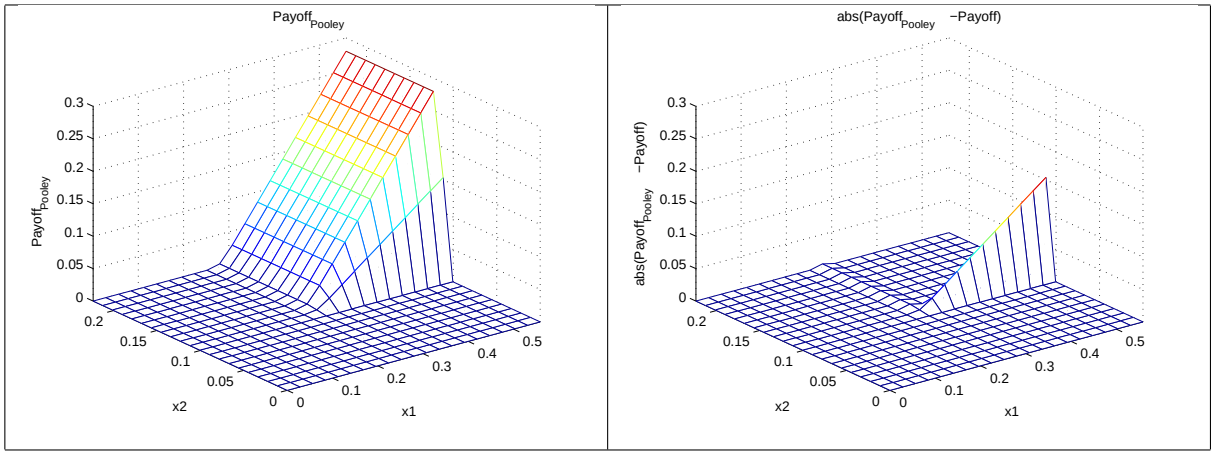

FIGURE 3.1.4. Example of localised effect around strikes with the Pooley method

3.1.6.1. *Localisation of the Pooley Method.* Figure (3.1.4) shows the plots for a two-dimensional payoff function that has been treated with the Pooley method. The payoff function is defined as the product of a call in one dimension with a strike 0.28 and a digital in the other dimension with a strike at 0.11, $\max(x_1 - 0.28, 0)\,\text{digital}(x_2 - 0.11)$. On the left is the payoff treated by the Pooley method; on the right is the absolute difference between the Pooley-treated payoff and the payoff. After the Pooley treatment, it can be seen that this payoff changes little from its original function. Only the values around the

strikes have been modified. All other places remain unchanged. Therefore, the Pooley treatment can be applied to a very limited number of grid points, making the computation very economical.

3.1.6.2. *Separation of Dimensions of the Pooley Method.* For a n-dimensional space, (3.1.17) can be written as

$$F_k = \left(\prod_{i=1}^{n} \delta x_i^{-1}\right) \int \int ... \int N_k(\boldsymbol{x}) F(\boldsymbol{x}) dx_1 dx_2 ... dx_n \tag{3.1.19}$$

where x_i is the element of the vector $\boldsymbol{x}$.

We can choose to express $F(\boldsymbol{x})$ and $N_k(\boldsymbol{x})$ using a product of piecewise linear functions in each dimension.

$$\begin{aligned} F(\boldsymbol{x}) &= \textstyle\prod_{i=1}^{n} F^{(i)}(x_i) \\ N_k(\boldsymbol{x}) &= \textstyle\prod_{i=1}^{n} N_k^{(i)}(x_i) \end{aligned} \tag{3.1.20}$$

Combining (3.1.20) and (3.1.17) gives

$$F_k = \prod_{i=1}^{n} \delta x_i^{-1} \int N_k^{(i)}(x_i) F^{(i)}(x_i) dx_i \tag{3.1.21}$$

Many payoff functions allow us to reduce the multi-dimensional integral to a product of one-dimensional integrals. But there are a few exceptions like $\max(x_1 - x_2 - K)$ and $\text{digital}(x_1 - x_2 - K)$. These functions will need to involve embedded loops to complete the numerical computation.

3.1.7. Computing Sensitivities through "Bumping". Backward time marching automatically yields Δ and Γ along with V. When we know the discretised solution V_j, where $j \in [1, 2... J]$, we can compute Δ_j and Γ_j by central differencing with V_{j-1}, V_j, and V_{j+1}. Some sensitivities such as Vega and Vanna can not be obtained this easily. For some models and instruments, analytical solutions exist for the common Greeks.

For those sensitivities that have to be obtained numerically, a common industrial practice is through parameter perturbation, an easy-to-implement method also called "bumping". For example, the value of $\dfrac{\partial^2 V}{\partial \sigma^2}$ can be computed through the values of $V(\sigma + \Delta\sigma)$, $V(\sigma)$ and $V(\sigma - \Delta\sigma)$, where the perturbed $\Delta\sigma$ value is chosen to be small.

3.2. Convergence Results for Second-order Schemes

3.2.1. Giles-Carter Results on Non-smooth Payoffs with the Black-Scholes model. This section is meant to verify and present the results made by Giles & Carter [37] in summary.

We use the Black-Scholes model for our numerical tests because the Black-Scholes model has nice analytical formula for the price and the Greeks, and the convergence results come from comparing the analytical and numerical solutions. The Black-Scholes equation[1] was derived by Fischer Black and Myron Scholes [8] in 1973. The underlying asset is assumed to follow geometric Brownian motion.

$$dS = \mu S dt + \sigma S dX \tag{3.2.1}$$

which leads to the PDE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r\, S \frac{\partial V}{\partial S} - rV = 0 \tag{3.2.2}$$

For European call and put options, explicit formulas exist for the solution of Black-Scholes equation.

Assume the terminal condition $(S-K,0)^+$ for a European call. The solution is

$$C(S,t) = S\mathcal{N}(d_1) - Ke^{-r(T-t)}\mathcal{N}(d_2) \tag{3.2.3}$$

where

$$\mathcal{N}(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{1}{2}y^2} dy \tag{3.2.4}$$

$$d_1 = \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma(\sqrt{T-t})} \tag{3.2.5}$$

$$d_2 = \frac{\log(S/K) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma(\sqrt{T-t})} \tag{3.2.6}$$

The terminal condition for a European put is $(K-S,0)^+$. The solution for a put option is

$$P(S,t) = -S\mathcal{N}(-d_1) + Ke^{-r(T-t)}\mathcal{N}(-d_1) \tag{3.2.7}$$

By taking first and second derivatives of option price with respect to the underlying, we obtain Δ and Γ. Δ measures the sensitivity of the option to the underlying while Γ

[1] See [52], [101], [99], [51] for related topics in this section.

measures the sensitivity of the delta to the underlying.

$$\Delta = \frac{\partial V}{\partial S} \tag{3.2.8}$$

$$\Gamma = \frac{\partial^2 V}{\partial S^2} \tag{3.2.9}$$

For European options with no dividends, the Greeks for the Black-Scholes model have the following formulas:

$$\Delta_{call} = \mathcal{N}(d_1) \tag{3.2.10}$$

$$\Delta_{put} = \mathcal{N}(d_1) - 1 \tag{3.2.11}$$

$$\Gamma_{call} = \frac{\mathcal{N}'(d_1)}{\sigma S\sqrt{T-t}} \tag{3.2.12}$$

$$\Gamma_{put} = \frac{\mathcal{N}'(d_1)}{\sigma S\sqrt{T-t}} \tag{3.2.13}$$

The payoff function is not smooth around the strike. As pointed out earlier, the option price V and its first-order derivative $\frac{\partial V}{\partial S}$ have limited local $\mathcal{L}^2$ norm but $\frac{\partial^2 V}{\partial S^2}$ does not. It means that over a limited length of discretised domain, we have

$$\lim_{J\to+\infty} \sqrt{\sum_{j=1}^{J} (V_j)^2 \, \delta S} < +\infty \tag{3.2.14}$$

$$\lim_{J\to+\infty} \sqrt{\sum_{j=1}^{J} (\frac{\partial V_j}{\partial S})^2 \delta S} < +\infty \tag{3.2.15}$$

$$\lim_{J\to+\infty} \sqrt{\sum_{j=1}^{J} (\frac{\partial^2 V_j}{\partial S^2})^2 \delta S} = +\infty \tag{3.2.16}$$

where j denotes the index of grid point, the total number of grid points is J, δS is the grid interval of S and $\frac{\partial V_j}{\partial S}$, $\frac{\partial^2 V_j}{\partial S^2}$ are the discretised delta and gamma at the point j.

The plots in this section exhibit the contrast between the numerical and analytical solutions under regular Crank-Nicolson method and Crank-Nicolson method with Rannacher start-up for a European put option. The strike is set to 1. The computation is conducted over S between 0 and 4, r set to 0.02, the terminal time T set to 1, and $\sigma = 0.14$. In the plots, μ denotes $\sigma^2 K \frac{\delta t}{\delta S}$, where δt is the time interval.

In Figure (3.2.1) we see a good match of numerical and analytical solutions for V using the Crank-Nicolson method. In the non-smooth payoff area, there is no apparent non-smoothness in V. For this particular μ, a small area of discontinuity sets in around the

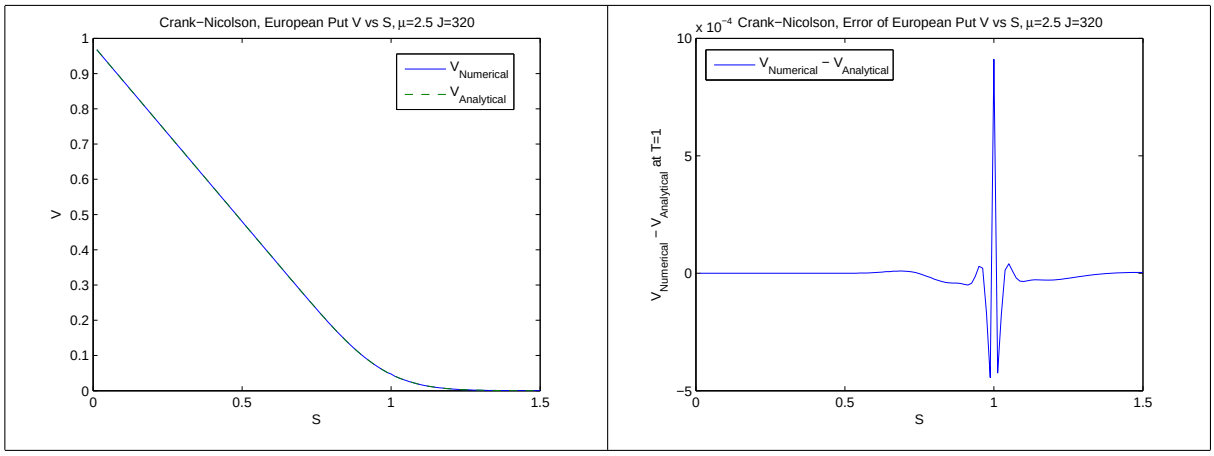

FIGURE 3.2.1. Crank-Nicolson, V, European Put

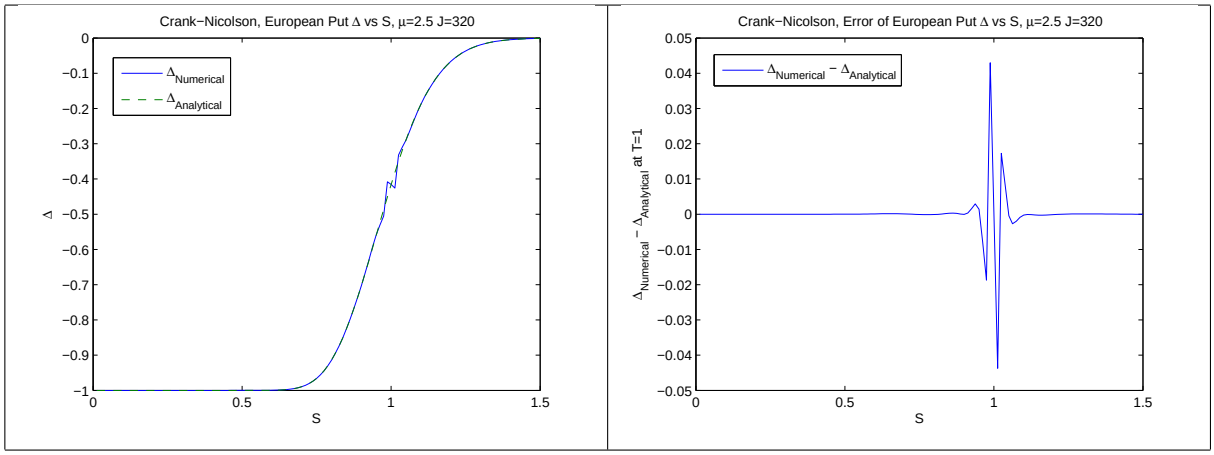

FIGURE 3.2.2. Oscillation in Crank-Nicolson, $\Delta = V_S$, European Put

strike in the plot for Δ in Figure (3.2.2). For this particular μ, a huge oscillation appears around the strike in the plot for Γ in Figure (3.2.3). This oscillation as we shall later see in the convergence plots will not go away if we refine the grid. Rather, it will grow even bigger.

The initial smoothing introduced by the Rannacher start-up allows the Crank-Nicolson method to overcome the $\mathcal{L}^2$ norm restriction. Thus in Figures (3.2.4), (3.2.5) and (3.2.6) a good fit between solutions is illustrated.

Figure (3.2.7) shows that for smaller μ, V converges in second-order. However, as μ grows, an "elbow" appears and the convergence line becomes first-order in the part beyond the "elbow" point.

Figure (3.2.8) shows that for smaller μ, Δ apparently converges in second-order. However, as μ grows, an "elbow" appears and the convergence line becomes flat (0th order) in the part beyond the "elbow" point.

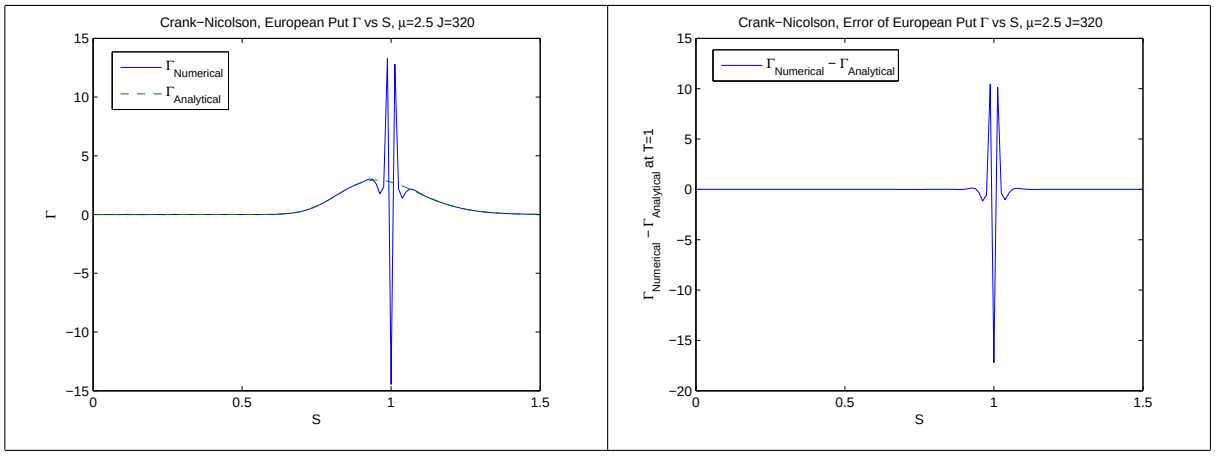

FIGURE 3.2.3. Oscillation in Crank-Nicolson, $\Gamma = V_{SS}$, European Put

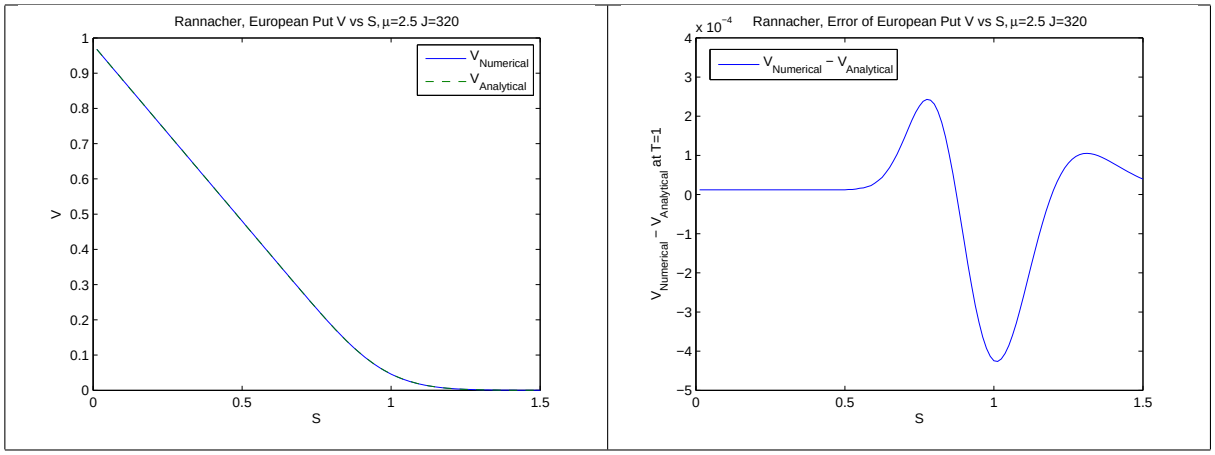

FIGURE 3.2.4. Rannacher Start-up, V, European Put

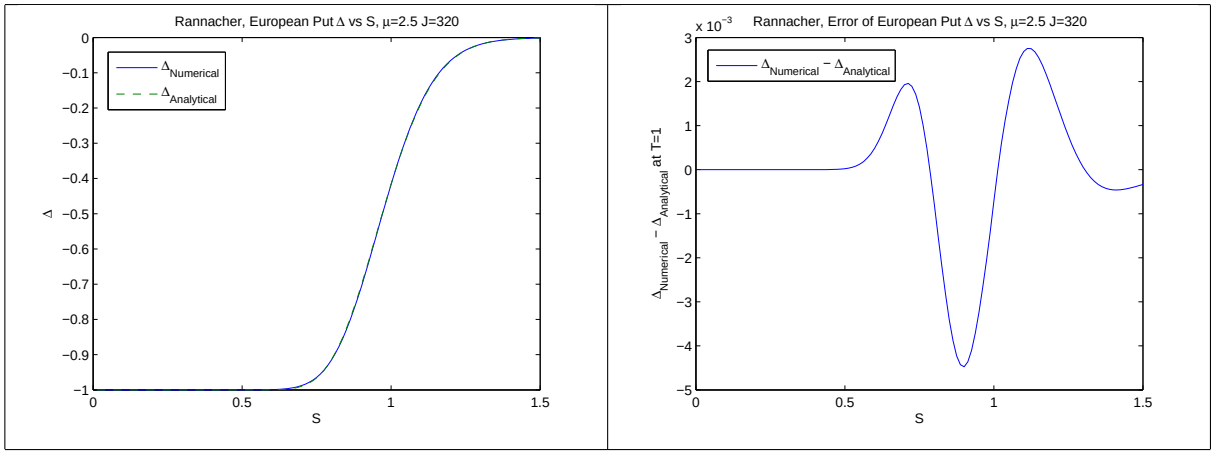

FIGURE 3.2.5. Rannacher Start-up, $\Delta = V_S$, European Put

Figure (3.2.9) shows that for smaller μ, Γ apparently converges in second-order. However, as μ grows, an "elbow" appears and the line becomes first-order divergent in the part beyond the "elbow" point.

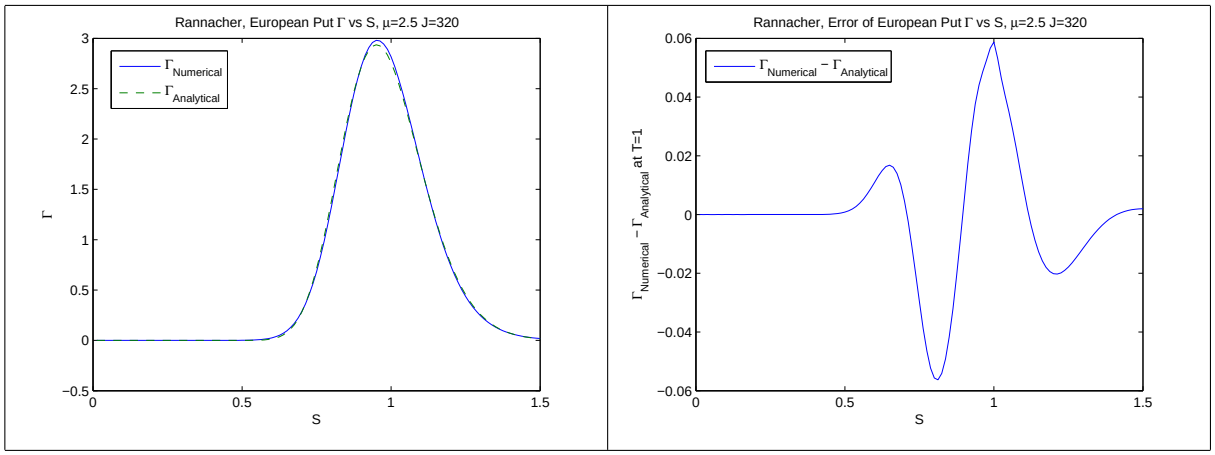

FIGURE 3.2.6. Rannacher Start-up, $\Gamma = V_{SS}$, European Put

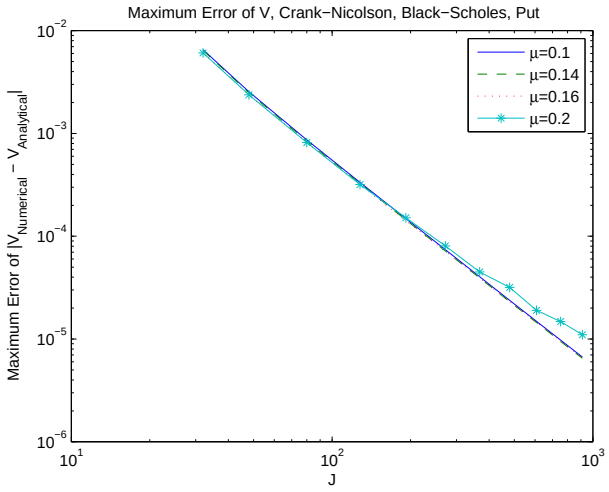

FIGURE 3.2.7. Deviation from second-order convergence of Crank-Nicolson, V, European Put

With the Rannacher start-up, a restoration of second-order convergence is evident in Figure (3.2.10), Figure (3.2.11) and Figure (3.2.12).

Giles et al [37] show that the Rannacher start-up works for non-smooth data in the $\mathcal{L}^\infty$ norm. Although the Rannacher start-up is useful in restoring second-order convergence for nonsmooth payoffs, the Rannacher start-up introduces a little loss of accuracy compared to the standard Crank-Nicolson method for smooth payoffs. That is because the initial steps for the Rannacher start-up use the Backward Euler method and are only first-order.

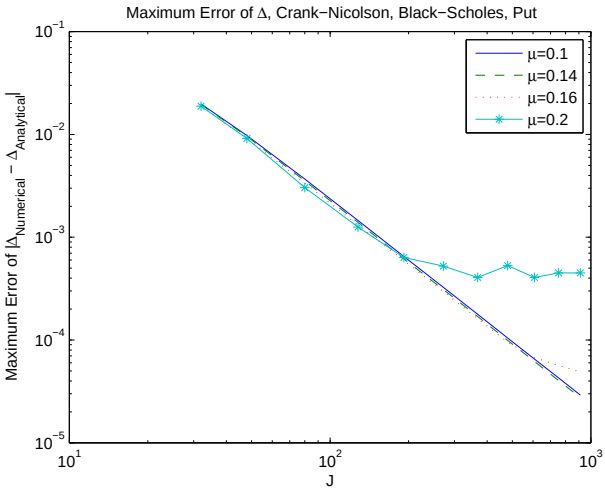

FIGURE 3.2.8. Flat error of Crank-Nicolson, $\Delta = V_S$, European Put

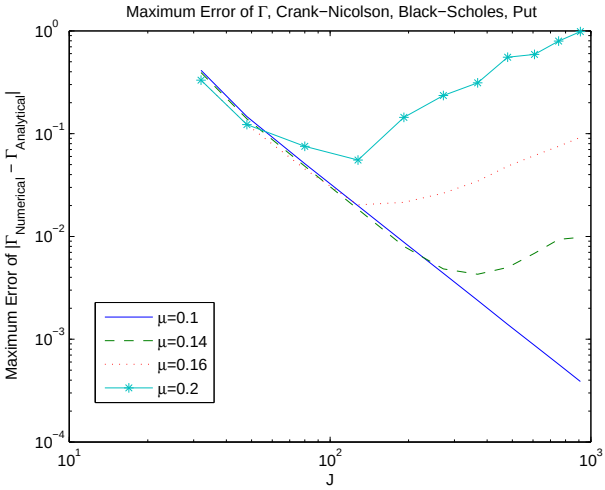

FIGURE 3.2.9. First-order error divergence of Crank-Nicolson, $\Gamma = V_{SS}$, European Put

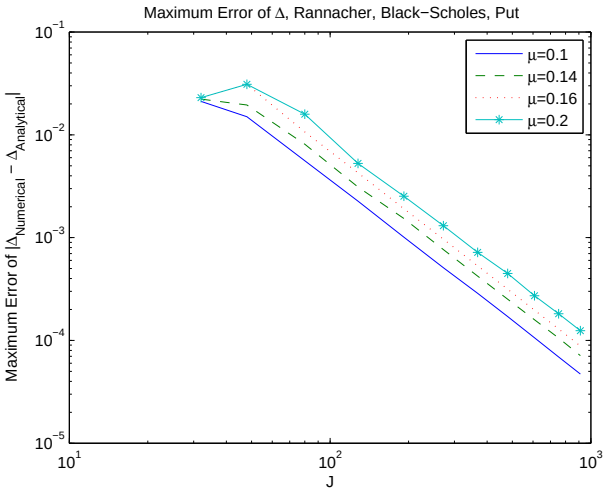

FIGURE 3.2.11. Second order convergence of Rannacher Start-up, $\Delta = V_S$, European Put

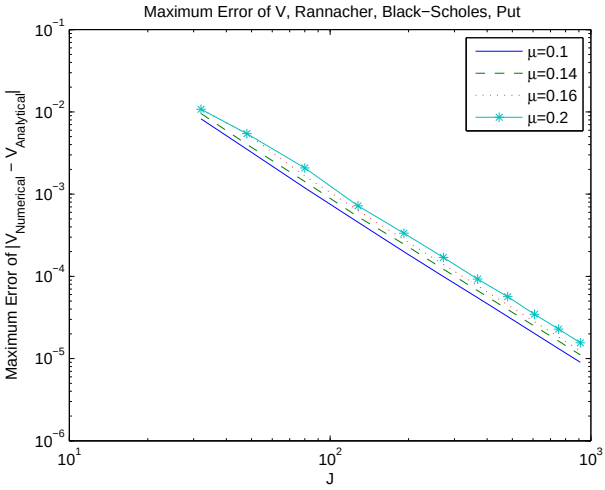

FIGURE 3.2.10. Second order convergence of Rannacher Start-up, V, European Put

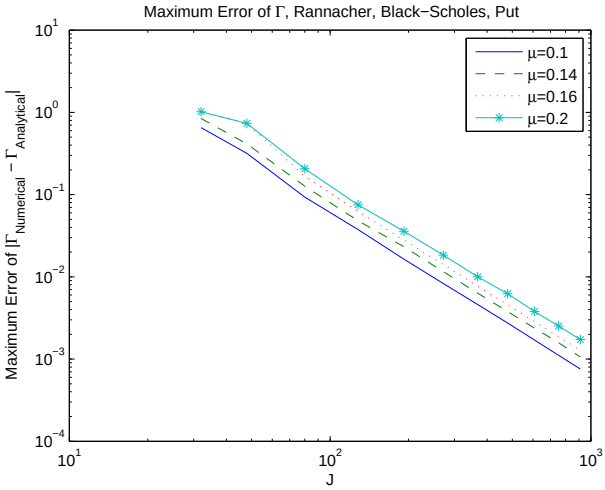

FIGURE 3.2.12. Second order convergence of Rannacher Start-up, $\Gamma = V_{SS}$, European Put

3.2.2. Multi-dimensional Numerical Convergence. We will use the short rate model to demonstrate the performance of the second order Craig-Sneyd scheme and the performance of the co-ordinate transform method when the co-ordinate transform can be used to eliminate the cross-derivative term. The particular form of the PDE and its analytical solution will be introduced, followed by the numerical results.

The spot interest rate r in certain models can be expressed by the following stochastic differential equation.[2]

[2]See [80], [91], [6] for related topics in this section.

$$dr = \mu(t,r)dt + \sigma(t,r)dW \tag{3.2.17}$$

The methodology by Vasicek [96] gives the partial differential equation for the price of a zero-coupon bond

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} + (\mu - \lambda\sigma)\frac{\partial V}{\partial r} - rV = 0 \tag{3.2.18}$$

where $\lambda(r,t)$ is called market price of risk.

The class of normal models describe the spot interest rate as follows (80):

$$\begin{aligned} r &= F(t,r^*) \\ dr^* &= (\theta^*(t) - a^*(t)r^*)dt + \sigma(t)dW \end{aligned} \tag{3.2.19}$$

Due to the fact the process r^* is linear, r^* has a normal distribution. $F(t,r^*) = r^*$ gives the Hull and White model [49]. When $F(t,r^*) = (r^*)^2$, (3.2.19) becomes the squared Gaussian model. If $F(t,r^*)$ equals e^{r^*}, (3.2.19) becomes the Black-Karasinski model [7]. It is the Ho-Lee model [46] for the conditions $a^*(t) \equiv 0$, $\sigma(t) \equiv \sigma$, and $F(t,r) \equiv r^*$.

Using the Hull-White model gives

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} + ((\theta^*(t) - a^*(t)r) - \lambda(t,r)\,\sigma(t))\,\frac{\partial V}{\partial r} - rV = 0 \tag{3.2.20}$$

Assuming that the market price of risk $\lambda(t,r)$ takes the form of $\lambda_1(t) + \lambda_2(t)r$ (49, 80), (3.2.20) becomes

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial r^2} + (\theta(t) - a(t)r)\frac{\partial V}{\partial r} - rV = 0 \tag{3.2.21}$$

where

$$\begin{aligned} \theta(t) &= \theta^*(t) - \lambda_1(t)\sigma(t) \\ a(t) &= a^*(t) + \lambda_2(t)\sigma(t) \end{aligned} \tag{3.2.22}$$

This Hull-White model can be extended by combining a Heston-type stochastic volatility. The extended process for r is assumed to be

$$\begin{aligned} dr &= (\theta_r^*(t) - a_r^*(t)r)dt + \sigma_r(t)\sqrt{\nu}dW_1 \\ d\nu &= (\theta_\nu(t) - a_\nu(t)\nu)dt + \sigma_\nu(t)\sqrt{\nu}dW_2 \end{aligned} \tag{3.2.23}$$

where $dW_1 dW_2 = \rho dt$. In this model, r has a mean reverting stochastic volatility expressed by $\sqrt{\nu}$.

Similarly, the PDE for this model is obtained as

$$\frac{\partial V}{\partial t}+(\theta_r-a_r r)\frac{\partial V}{\partial r}+(\theta_\nu-a_\nu\nu)\frac{\partial V}{\partial \nu}+\frac{1}{2}\sigma_r^2\,\nu\frac{\partial^2 V}{\partial r^2}+\frac{1}{2}\sigma_\nu^2\,\nu\frac{\partial^2 V}{\partial \nu^2}+\rho\sigma_r\sigma_\nu\nu\frac{\partial^2 V}{\partial r\partial \nu}=rV. \quad (3.2.24)$$

where $\theta_r(t)=\theta_r^*(t)-\lambda_1(t)\sigma_r(t)$ and $a_r(t)=a_r^*(t)+\lambda_2(t)\sigma_r(t)$. A zero-coupon bond has the unit terminal payoff $V(T)=1$, where T is the maturity.

The semi-analytical solution for the above equation for a zero-coupon bond is given by

$$V(r,\nu,t)=\exp\left(b_1(T-t)+r\,b_2(T-t)+\nu\,b_3(T-t)\right) \quad (3.2.25)$$

where $b_1(t)$, $b_2(t)$, $b_3(t)$ satisfies the o.d.e system

$$\frac{d}{dt}\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}=\begin{pmatrix} \theta_r b_2+\theta_\nu b_3 \\ -a_r b_2-1 \\ -a_\nu b_3+\frac{1}{2}\sigma_r^2 b_2^2+\rho\sigma_r\sigma_\nu b_2 b_3+\frac{1}{2}\sigma_\nu^2 b_3^2 \end{pmatrix} \quad (3.2.26)$$

subject to the initial conditions $b_1(0)=b_2(0)=b_3(0)=0$. As can be seen, the zero-coupon bond $V(r,\nu,t)$ pays off one at $t=T$.

Equation (3.2.24) was used for computation of convergence results. The following parameter values were assumed: $r=0.02$, $\nu=0.001$, $\theta_r=0.002$, $a_r=0.1$, $\sigma_r=1$, $\theta_\nu=0.0003$, $a_\nu=0.3$, $\sigma_\nu=0.003$, $\rho=0.5$, and $T=15$.

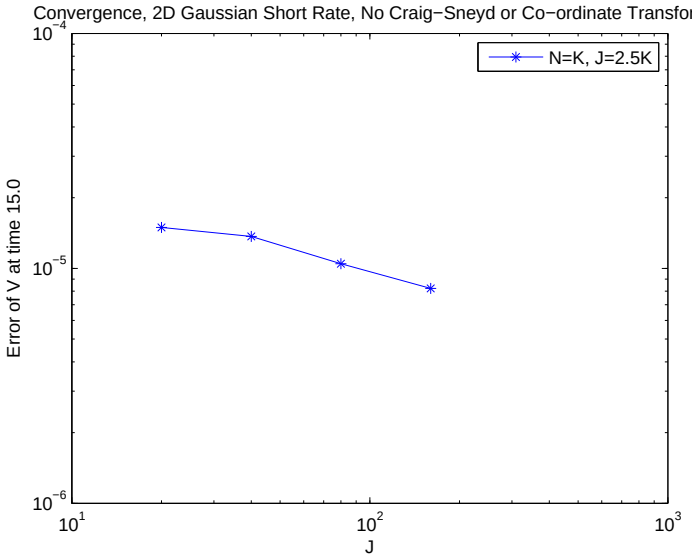

FIGURE 3.2.13. Error with only Crank-Nicolson, short rate model for zero-coupon bond

With the introduction of the cross-derivative term, the splitted Crank-Nicolson method fails to maintain second-order convergence as shown in Figure (3.2.13). The convergence rate is at best first-order in this case.

One solution is to use the Craig-Sneyd correction. As shown in Figure (3.2.14), the Craig-Sneyd correction well handles this term and nicely restores second-order convergence. Another approach is to use transformation of co-ordinates with the resulting convergence plot shown in Figure (3.2.15).

The co-ordinate transform method in this particular example restores second-order convergence as well. A smooth co-ordinate transform does not usually bring qualitative changes to convergence performance. When the ADI method is used and the cross-derivative term is present, however, co-ordinate transform that eliminates the cross-derivative term may improve convergence performance. But not all problems have the desired second-order effect with co-ordinate transform. If co-ordinate transform is not properly applied to smooth out the discretised payoff, a second-order convergence will not be seen.

Compared to transformation of co-ordinates approach, the Craig-Sneyd method is easier to implement. With it, the effort of deriving a proper co-ordinate transformation is spared. The code for the standard Crank-Nicolson scheme can be kept intact for the most part. Essentially, the Craig-Sneyd method is similar to running the standard Crank-Nicolson scheme twice. Thus very little code modification is involved when the Craig-Sneyd correction is needed on top of the Crank-Nicolson method.

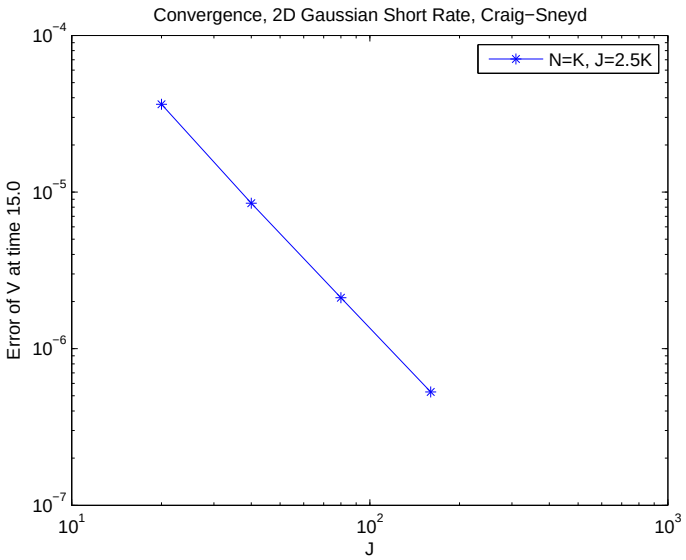

FIGURE 3.2.14. Second order error convergence of Craig-Sneyd on short rate model for zero-coupon bond

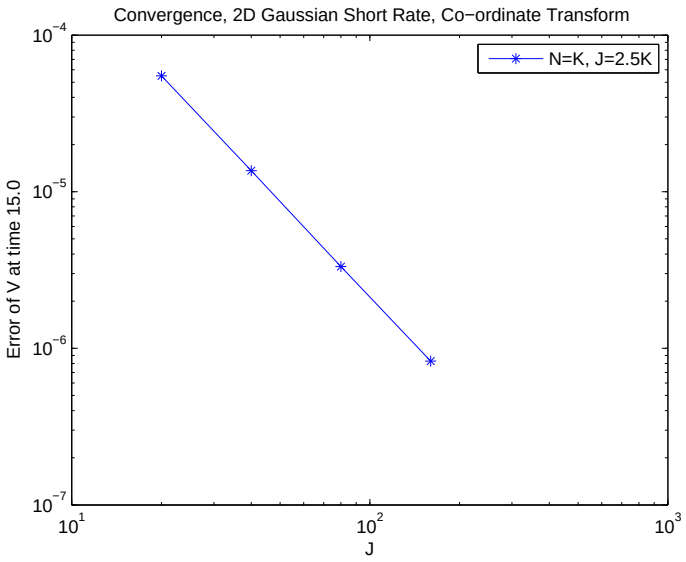

FIGURE 3.2.15. Second order error convergence of co-ordinate transform on short rate model for zero-coupon bond

3.2.3. Payoff Smoothing by the Pooley and Averaging Methods. We will demonstrate the error performance by the Pooley and averaging methods using the Ornstein-Uhlenbeck (OU) process.

Assume the underlying follows a d-dimensional OU process:[3]

$$dx_i = -\lambda_i x_i dt + \sigma_i dW_i \tag{3.2.27}$$

where $1 \leqq i \leqq d$, $\lambda_i > 0$, $\sigma_i > 0$, λ_i and σ_i are nonrandom, measurable, and locally bounded; W_i is a Wiener process, $dW_i(t)dW_j(t) = \rho_{ij}(t)dt$, $1 \leqq i \leqq d$.

Let $\mathbf{x}$ denote the vector $[x_1, x_2, \ldots x_d]^T$. According to the Feynman-Kac theorem, the value function $E_t[F(\mathbf{x})]$ satisfies the following PDE:

$$\frac{\partial V}{\partial t} - \sum_i \lambda_i x_i \frac{\partial V}{\partial x_i} + \frac{1}{2}\sum_{i,j} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2 V}{\partial x_i \partial x_j} = 0 \tag{3.2.28}$$

subject to the terminal condition $V(\mathbf{x}, T) = F(\mathbf{x})$.

Given constant ρ_{ij}, σ_i and λ_i, the probability density function $p(\mathbf{x}, t)$ satisfies the adjoint PDE

$$-\frac{\partial p}{\partial t} + \sum_i \lambda_i \frac{\partial (p x_i)}{\partial x_i} + \frac{1}{2}\sum_{i,j} \rho_{ij}\sigma_i\sigma_j \frac{\partial^2 p}{\partial x_i \partial x_j} = 0 \tag{3.2.29}$$

subject to the initial condition $p(\mathbf{x}, 0) = \delta(\mathbf{x} - \mathbf{x}_0)$.

[3]See [77], [56], [76] for related topics in this section.

The above equation has the solution [34]

$$p(\mathbf{x}_0;\mathbf{x},t)=\frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}}\exp(-\frac{1}{2}(\mathbf{x}-e^{-\Lambda t}\mathbf{x}_0)^T\Sigma^{-1}(\mathbf{x}-e^{-\Lambda t}\mathbf{x}_0)) \tag{3.2.30}$$

where $\Lambda_{ij,\,i=j}=\lambda_i$ and $\Lambda_{ij,\,i\neq j}=0$, and the covariance matrix satisfies

$$\Sigma_{i,j}(t)=\frac{\rho_{ij}\sigma_i\sigma_j}{\lambda_i+\lambda_j}(1-\exp(-(\lambda_i+\lambda_j)t))). \tag{3.2.31}$$

Figure (3.2.16) illustrates the error reduction effect with the Pooley and averaging method. The numerical test is done by applying Crank-Nicolson scheme and Rannacher Start-up to one dimensional mean reverting PDE (3.2.28) with the parameters $\lambda=0.05$, $T=5$, $K=0.2795$ and $\sigma=0.03$. The payoff is a digital function. The grid is adjusted so that the strike K falls between the two grid points x_{j-1} and x_j. In the first plot, the strike is $0.15\,\delta x$ away from a grid point and in the second plot, the strike is $0.4\,\delta x$ away from a grid point, where δx denotes the grid interval.

The one dimensional model is used because we want to show the error effect of the strike location within the grid interval clearly. For higher dimensional models, the Pooley method is applied using (3.1.19), or (3.1.21) if possible, for which Figure (3.1.4) gives an example. The Rannacher Start-up can also be easily applied to multidimensional models with ADI. Take a two-dimensional model for example: the discretisation for the fully implicit scheme in ADI can be expressed by $(1-D_x)\,(1-D_y)\,V^{n+1}=V^n$; what the Rannacher scheme needs is just to run two half timesteps of the fully implicit scheme.

The performance of the Pooley method is better than that of the averaging method. In particular, when the strike is closer to a grid point (as in the first plot), the improvement appears more pronounced.

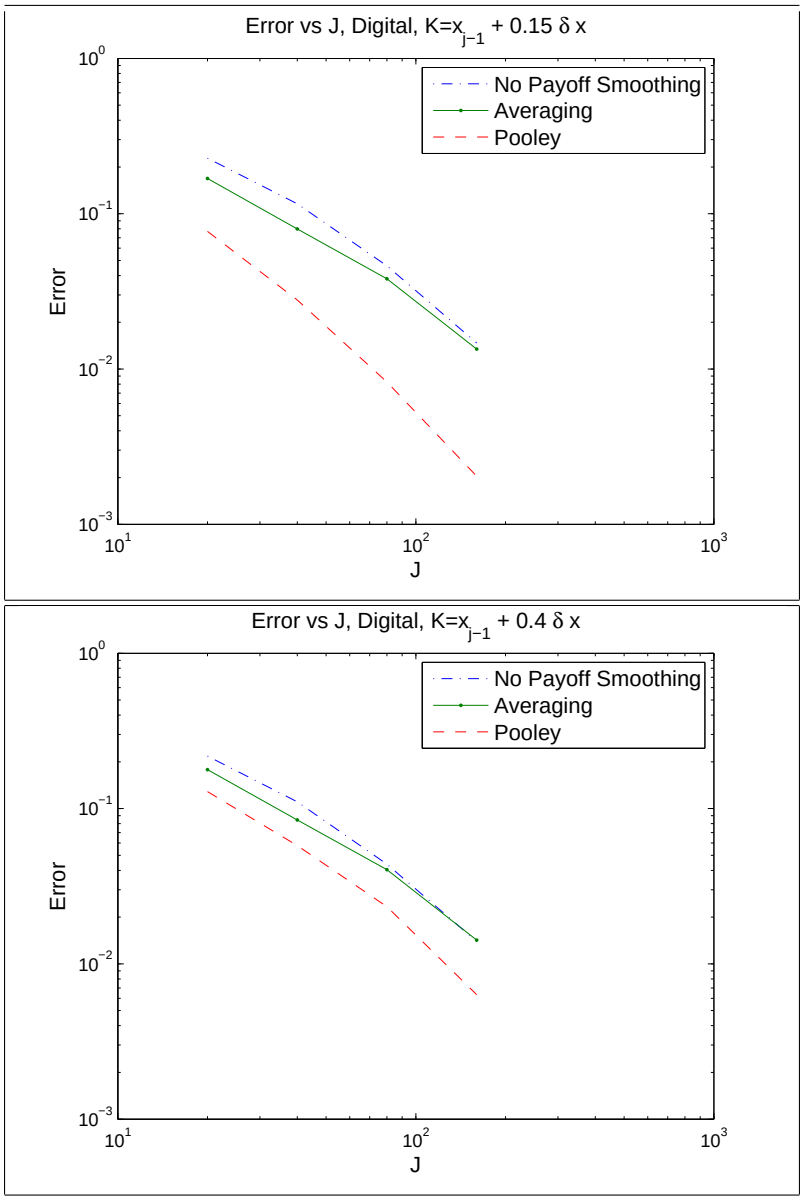

FIGURE 3.2.16. Error improvement with the Pooley and averaging methods

CHAPTER 4

Multi-Heston Stochastic Model

4.1. Multi-Heston Model Problem

Kainth et al [54] proposed the Multi-Heston model, an extended Heston model with multiple stochastic variance components

$$\begin{aligned} \frac{dS_t}{S_t} &= (r-q)\,dt + \sum_{j=1}^{N} \sqrt{\nu_t^j}\, dW_t^j \\ d\nu_t^j &= \kappa_j \left(\theta_j - \nu_t^j\right) dt + \xi_j \sqrt{\nu_t^j}\, dZ_t^j \end{aligned} \tag{4.1.1}$$

where W_t^j, and Z_t^j are Wiener processes, and $j \in \{1, 2, ...N\}$; $dW_t^m dZ_t^m = \rho_m dt$ for $m \in \{1, 2, ...N\}$; $dW_t^m dZ_t^n = dW_t^m dW_t^n = dZ_t^m dZ_t^n = 0$ for $m \neq n$, $m, n \in \{1, 2, ...N\}$. Here r is the interest rate and q is the dividend rate. When the model is used in FX, r means the domestic interest rate r_d and q means the foreign interest rate r_f. The drift $r - q$ and model parameters κ_j, θ_j, and ξ_j are not limited to constants. However, for the case of these being constant, a pseudo-analytical solution exists. Christoffersen et al [20] proposed the Double-Heston model, essentially the Multi-Heston model (4.1.1) with two stochastic variance factors.

For constant parameters, Kainth et al [54] and Christoffersen et al [20] gave the analytical solutions for the Multi-Heston model and Double-Heston model, respectively. Other than the difference in number of dimensions, the solutions developed by the two group of authors are exactly the same. Hence the solution developed by Kainth et al [54] is more general.

This chapter will be mainly concerned with the constant-parameter Multi-Heston model.

The Multi-Heston model is not a sum of the Heston models of individual instruments in a portfolio. For example, assume we have to price an option, of which the underlying is the sum of N instruments. Each one of the underlying instruments is assumed to satisfy

a Heston process

$$\begin{aligned} dS_t^j &= (r-q_j)\,S_t^j dt + S_t^j\sqrt{\nu_t^j}dW_t^j \\ d\nu_t^j &= \kappa_j\left(\theta_j-\nu_t^j\right)dt + \xi_j\sqrt{\nu_t^j}dZ_t^j \end{aligned} \tag{4.1.2}$$

The sum of the instruments, $\Pi_t = \sum_{j=1}^N S_t^j$, will follow

$$\begin{aligned} d\Pi_t &= \sum_{j=1}^N (r-q_j)\,S_t^j dt + \sum_{j=1}^N S_t^j\sqrt{\nu_t^j}dW_t^j \\ d\nu_t^j &= \kappa_j\left(\theta_j-\nu_t^j\right)dt + \xi_j\sqrt{\nu_t^j}dZ_t^j \end{aligned} \tag{4.1.3}$$

Obviously,

$$\begin{aligned} d\Pi_t &= \sum_{j=1}^N (r-q_j)\,S_t^j dt + \sum_{j=1}^N S_t^j\sqrt{\nu_t^j}dW_t^j \\ &\neq (r-q)\left(\sum_{j=1}^N S_t^j\right)dt + \left(\sum_{j=1}^N S_t^j\right)\sum_{j=1}^N\sqrt{\nu_t^j}dW_t^j \end{aligned}$$

where q is assumed to be the dividend for Π_t. So Π_t does not follow the Multi-Heston model.

Although the Multi-Heston underlying is not a simple sum of Heston underlyings, it is close to a geometric mean of the Heston underlyings but not in an exact way. Re-arrange (4.1.2) into the log form and sum them up to become

$$\begin{aligned} d\left(\sum_{j=1}^N \ln S_t^j\right) &= \sum_{j=1}^N\left(r-q_j-\frac{1}{2}\nu_t^j\right)dt + \sum_{j=1}^N\sqrt{\nu_t^j}dW_t^j \\ d\nu_t^j &= \kappa_j\left(\theta_j-\nu_t^j\right)dt + \xi_j\sqrt{\nu_t^j}dZ_t^j \end{aligned}$$

The analytical solution for a call option given by Kainth et al [54] is

$$V_{call} = P_1(S,\nu,\tau) - Ke^{-(r-q)\tau}P_2(S,\nu,\tau) \tag{4.1.4}$$

where $F = Se^{(r-q)\tau}$, the time to maturity $\tau = T-t$,

$$\begin{cases} P_1 &= \dfrac{1}{2}+\dfrac{1}{\pi}\displaystyle\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\phi\ln K}\varphi_t(\phi-i)}{i\phi F}\right)d\phi, \\ P_2 &= \dfrac{1}{2}+\dfrac{1}{\pi}\displaystyle\int_0^\infty \mathrm{Re}\left(\frac{e^{-i\phi\ln K}\varphi_t(\phi)}{i\phi}\right)d\phi, \end{cases} \tag{4.1.5}$$

and

$$\begin{aligned}
\varphi(S,\nu_j,\tau;\phi) &= \exp\left[\sum_{j=1}^{N}\left(A_j^*(\tau,\phi)+B_j^*(\tau,\phi)\nu_j\right)+i\phi\ln F\right],\\
A_j^*(\tau,\phi) &= \frac{\kappa_j\theta_j}{\xi_j^2}\left(\left(\kappa_j-\rho_j\xi_j\phi i+d_j^*\right)\tau-2\ln\left(\frac{c_j^*e^{d_j^*\tau}-1}{c_j^*-1}\right)\right),\\
B_j^*(\tau,\phi) &= \frac{\kappa_j-\rho_j\xi_j\phi i+d_j^*}{\xi_j^2}\left(\frac{e^{d_j^*\tau}-1}{c_j^*e^{d_j^*\tau}-1}\right),\\
c_j^*(\phi) &= \frac{\kappa_j-\rho_j\xi_j\phi i+d_j^*}{\kappa_j-\rho_j\xi_j\phi i-d_j^*},\\
d_j^*(\phi) &= \sqrt{(\rho_j\xi_j\phi i-\kappa_j)^2+i\xi_j^2\phi+\xi_j^2\phi^2}.
\end{aligned}\tag{4.1.6}$$

4.1.1. Pseudo-Analytical Multi-Heston Solutions through Fundamental Transform. Lewis [65] and Shaw [89] used an elegant fundamental transform technique to derive the pseudo-analytical solution for the Heston model. The same process of derivation can be applied to the Multi-Heston model.

The Multi-Heston model through the Feynman-Kac theorem gives the following Multi-Heston PDE

$$\begin{aligned}
-\frac{\partial V}{\partial\tau}+(r-q)\,S\frac{\partial V}{\partial S}+\frac{1}{2}\left(\sum_{j=1}^{N}\nu_j\right)S^2\frac{\partial^2V}{\partial S^2}+\sum_{j=1}^{N}\kappa_j(\theta_j-\nu_j)\frac{\partial V}{\partial\nu_j}\\
+\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\nu_j\frac{\partial^2V}{\partial\nu_j^2}+S\sum_{j=1}^{N}\xi_j\nu_j\rho_j\frac{\partial^2V}{\partial S\partial\nu_j}=rV
\end{aligned}\tag{4.1.7}$$

Applying the following co-ordinate transform

$$\begin{aligned}
x &= \ln(S)+(r-q)\,\tau\\
V(x,\,\nu_j,\,\tau) &= W(x,\,\nu_j,\,\tau)\exp(-r\tau)
\end{aligned}\tag{4.1.8}$$

to (4.1.7) gives

$$\begin{aligned}
\frac{1}{2}\sum_{j=1}^{N}\nu_j\left(\frac{\partial^2W}{\partial x^2}-\frac{\partial W}{\partial x}\right)+\sum_{j=1}^{N}\xi_j\nu_j\rho_j\frac{\partial^2W}{\partial x\partial\nu_j}\\
+\sum_{j=1}^{N}\kappa_j(\theta_j-\nu_j)\frac{\partial W}{\partial\nu_j}+\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\nu_j\frac{\partial^2W}{\partial\nu_j^2}=\frac{\partial W}{\partial\tau}
\end{aligned}\tag{4.1.9}$$

Define the Fourier Transform

$$\begin{aligned} W\left(x,\,\nu_j,\,\tau\right) &= \frac{1}{2\pi}\int_{i\phi_I-\infty}^{i\phi_I+\infty} e^{-i\phi x}\tilde{W}\left(\phi,\,\nu_j,\,\tau\right)d\phi \\ \tilde{W}\left(\phi,\,\nu_j,\,\tau\right) &= \int_{-\infty}^{+\infty} e^{i\phi x}W\left(x,\,\nu_j,\,\tau\right)dx \end{aligned} \tag{4.1.10}$$

where ϕ is a complex variable and $i\phi_I$ is its imaginary part.

Therefore, the Fourier transformed PDE for (4.1.9) becomes

$$\begin{gathered} \frac{1}{2}\sum_{j=1}^{N}\xi_j^2\nu_j\frac{\partial^2\tilde{W}}{\partial\nu_j^2}+\sum_{j=1}^{N}\left(\kappa_j(\theta_j-\nu_j)-i\phi\xi_j\nu_j\rho_j\right)\frac{\partial\tilde{W}}{\partial\nu_j} \\ \frac{1}{2}\left(\sum_{j=1}^{N}\nu_j\right)\left(i\phi-\phi^2\right)\tilde{W}=\frac{\partial\tilde{W}}{\partial\tau} \end{gathered} \tag{4.1.11}$$

At the maturity $\tau = 0$,

$$\tilde{W}\left(\phi,\,\nu_j,\,0\right)=\int_{-\infty}^{\infty}e^{i\phi x}W\left(x,\,\nu_j,\,0\right)dx=\int_{-\infty}^{\infty}e^{i\phi x}V\left(x,\,\nu_j,\,0\right)dx$$

For the call, put, digital call and digital put payoffs, Shaw [89] gave $\tilde{W}\left(\phi,\,\nu_j,\,0\right)$ for different payoffs in the following table:

Payoffs	$V\left(x,\,\nu_j,\,0\right)$	$\tilde{W}\left(\phi,\,\nu_j,\,0\right)$	Restriction
Call	$\max\left(e^x-K,\,0\right)$	$\dfrac{K^{(1+i\phi)}}{i\phi-\phi^2}$	$\phi_I=\mathrm{Im}\left(\phi\right)>1$
Put	$\max\left(K-e^x,\,0\right)$	$\dfrac{K^{(1+i\phi)}}{i\phi-\phi^2}$	$\phi_I=\mathrm{Im}\left(\phi\right)<0$
Digital Call	$\begin{cases} 1, & e^x>K \\ 0, & K<e^x \end{cases}$	$-\dfrac{K^{i\phi}}{i\phi}$	$\phi_I=\mathrm{Im}\left(\phi\right)>0$
Digital Put	$\begin{cases} 0, & e^x>K \\ 1, & K<e^x \end{cases}$	$\dfrac{K^{i\phi}}{i\phi}$	$\phi_I=\mathrm{Im}\left(\phi\right)<0$

TABLE 4.1. $\tilde{W}\left(\phi,\,\nu_j,\,0\right)$ for Different Payoffs

The "Restriction" columns give the conditions that allow the Fourier Transform integrals to converge. From these "Restriction" requirements, it can be seen that ϕ_I can be any constant that satisfies the corresponding "Restriction" for an option. For example, Lewis [65] suggested that a valid choice of ϕ_I for the call option will be 3/2, for it is larger than 1. For the above payoff functions or any payoff independent of volatility, $\tilde{W}\left(\phi,\,\nu_j,\,0\right)$ can be written as $\tilde{W}\left(\phi,\,0\right)$.

Assume a function $G\left(\phi,\,\nu_j,\,\tau\right)$ can be found so that

$$\tilde{W}\left(\phi,\,\nu_j,\,\tau\right)=\tilde{W}\left(\phi,\,0\right)G\left(\phi,\,\nu_j,\,\tau\right), \tag{4.1.12}$$

At $\tau = 0$, $G(\phi, \nu_j, 0) = 1$ obviously holds.

Because $\tilde{W}(\phi, 0)$ depends on neither τ nor ν_j, substituting (4.1.12) to (4.1.9) gives the ODE of $G(\phi, \nu_j, \tau)$ independent of $\tilde{W}(\phi, 0)$

$$\begin{gathered}\frac{1}{2}\sum_{j=1}^{N}\xi_j^2\nu_j\frac{\partial^2 G}{\partial\nu_j^2}+\sum_{j=1}^{N}\left(\kappa_j(\theta_j-\nu_j)-i\phi\xi_j\nu_j\rho_j\right)\frac{\partial G}{\partial\nu_j}\\ \frac{1}{2}\left(\sum_{j=1}^{N}\nu_j\right)\left(i\phi-\phi^2\right)G=\frac{\partial G}{\partial\tau}\end{gathered} \tag{4.1.13}$$

subject to $G(\phi, \nu_j, 0) = 1$.

Assume $G(\phi, \nu_j, \tau)$ takes the following form:

$$G(\phi, \nu_j, \tau) = \exp\left[\sum_{j=1}^{N}\left(A_j(\phi, \tau) + B_j(\phi, \tau)\nu_j\right)\right], \tag{4.1.14}$$

which is an extension to the form used by Heston [42]. Set $A_j(\phi, 0) = B_j(\phi, 0) = 0$ to satisfy the condition $G(\phi, \nu_j, 0) = 1$. Substituting (4.1.14) to (4.1.13) gives

$$\left[\frac{\partial A_j}{\partial\tau}-\kappa_j\theta_j B_j\right]+\nu\left[\frac{\partial B_j}{\partial\tau}-\frac{\xi_j^2B_j^2}{2}+(\kappa_j-i\phi\rho_j\xi_j)B_j-\frac{1}{2}i\phi+\frac{\phi^2}{2}\right]=0. \tag{4.1.15}$$

where $j \in \{1, ...N\}$.

To make (4.1.15) true for all ν, the following must hold:

$$\begin{cases}\dfrac{\partial A_j}{\partial\tau} &= \kappa_j\theta_j B_j,\\ \dfrac{\partial B_j}{\partial\tau} &= \dfrac{\xi_j^2B_j^2}{2}-(\kappa_j-i\phi\rho_j\xi_j)B_j+\dfrac{1}{2}i\phi-\dfrac{\phi^2}{2}.\end{cases} \tag{4.1.16}$$

subject to $A_j(\phi, 0) = B_j(\phi, 0) = 0$ for $j \in \{1, ...N\}$.

The ODEs involving B_j only are Riccati-type equations that have the following solutions:

$$B_j(\phi,\tau) = \frac{\kappa_j+\rho_j\xi_j\phi i+d_j}{\xi_j^2}\left(\frac{e^{d_j\tau}-1}{c_je^{d_j\tau}-1}\right) \tag{4.1.17}$$

where

$$\begin{cases}c_j(\phi) &= \dfrac{\kappa_j+\rho_j\xi_j\phi i+d_j}{\kappa_j+\rho_j\xi_j\phi i-d_j}\\ d_j(\phi) &= \sqrt{(\rho_j\xi_j\phi i+\kappa_j)^2-i\xi_j^2\phi+\xi_j^2\phi^2}\end{cases} \tag{4.1.18}$$

On knowing B_j, A_j can be obtained by integrating the first equation of (4.1.16):

$$A_j(\phi,\tau) = \sum_{j=1}^{N}\frac{\kappa_j\theta_j}{\xi_j^2}\left(\left(\kappa_j+\rho_j\xi_j\phi i+d_j\right)\tau-2\ln\left(\frac{c_je^{d_j\tau}-1}{c_j-1}\right)\right) \tag{4.1.19}$$

Lewis [65] called G a fundamental transform. Lewis [65] and Shaw [89] suggested the solution for a given payoff can be easily obtained by combining (4.1.8), (4.1.10), (4.1.12) and (4.1.14) into the general Multi-Heston option price

$$V\left(x,\,\nu_j,\,\tau\right)=\frac{e^{-r\tau}}{2\pi}\int_{i\phi_I-\infty}^{i\phi_I+\infty}e^{-i\phi x}\tilde{W}\left(\phi,\,0\right)\exp\left[\sum_{j=1}^{N}\left(A_j(\phi,\,\tau)+B_j(\phi,\,\tau)\nu_j\right)\right]d\phi \tag{4.1.20}$$

Notice that $A_j(\phi,\,\tau)=A_j^*(-\phi,\,\tau)$ and $B_j(\phi,\,\tau)=B_j^*(-\phi,\,\tau)$ where A_j^* and B_j^* follow (4.1.6). So (4.1.20) can be changed to

$$\begin{aligned}V &= \frac{e^{-r\tau}}{2\pi}\int_{-i\phi_I-\infty}^{-i\phi_I+\infty}e^{i\phi\ln F}\tilde{W}\left(-\phi,\,0\right)\exp\left[\sum_{j=1}^{N}\left(A_j^*(\phi,\,\tau)+B_j^*(\phi,\,\tau)\nu_j\right)\right]d\phi\\ &= \frac{e^{-r\tau}}{2\pi}\int_{-i\phi_I-\infty}^{-i\phi_I+\infty}\tilde{W}\left(-\phi,\,0\right)\varphi(\phi)d\phi\end{aligned} \tag{4.1.21}$$

4.2. Why Multiple Stochastic Variance Factors Matter

First of all, it should be noted that the Multi-Heston model is a one-factor model with respect to the underlying asset S_t. There is just one underlying S_t in the model; the multiple dW_t^j existing in the asset process only give rise to multiple variance processes. This becomes more evident in the Multi-Heston PDE that we will later see with only a single underlying variable.

Secondly, the Multi-Heston model can also be viewed as having a single variance factor. The variance of the return $\dfrac{dS}{S}$ in the Multi-Heston model (4.1.1) is computed as [20]

$$\mathrm{E}\left[\left(\frac{dS}{S}\right)^2\right]=\left(\sum_{j=1}^{N}\nu_t^j\right)dt.$$

The variance of the asset return is just the sum of the variances in the Multi-Heston model. Let $\nu_t=\left(\sum_{j=1}^{N}\nu_t^j\right)$. So the dS process in the Multi-Heston process can actually be written as

$$\frac{dS_t}{S_t}=\left(r-q\right)dt+\sqrt{\nu_t}dW_t=\left(r-q\right)dt+\sqrt{\sum_{j=1}^{N}\nu_t^j}dW_t$$

So the Multi-Heston process in fact has only one variance factor $\sqrt{\nu_t}dW_t$. All the individual variances ν_t^j influence the variance of the asset return through the sum of ν_t^j.

The sum of variances $d\nu_t$ process follows

$$d\nu_t = \sum_{j=1}^{N} \kappa_j \left(\theta_j - \nu_t^j\right) dt + \sum_{j=1}^{N} \xi_j \sqrt{\nu_t^j} dZ_t^j$$

The variance of $d\nu_t$ is

$$\mathrm{E}\left[(d\nu_t)^2\right] = \sum_{j=1}^{N} \xi_j^2 \nu_t^j$$

which is a weighted sum of ν_t^j.

If $d\nu_t$ could be written as a square root process $d\nu_t = \kappa(\theta - \nu_t)\, dt + \xi\sqrt{\nu_t} dZ_t^j$, the variance of $d\nu_t$ would have to be $\xi^2\nu_t$, not a weighted sum of ν_t^j. Also the mean reverting part $\sum_{j=1}^{N} \kappa_j \left(\theta_j - \nu_t^j\right) dt$ cannot be written as the form $\kappa(\theta - \nu_t)\, dt$ for a general case. This shows that the Multi-Heston model is not the same as the Heston model, hence it gives rise to possible properties that the Heston model does not have.

Define the log-moneyness $s = \ln(S/K)$ and log-strike $k = -s = \ln(K/S)$. Gatheral [36] gave the state price probability density for the Heston model by taking the sensitivity of P_2 in (4.1.5) with respect to s (or k) because P_2 represents Probability$(\ln S_T > \ln S)$. The same method of computing the probability density function $p(s)$ (or $p(k)$) can be applied to the Multi-Heston model:

$$p(s) = \frac{\partial P_2}{\partial s} = \frac{1}{\pi} \int_0^{\infty} \mathrm{Re}\left(e^{i\phi s} \exp\left[\sum_{j=1}^{N} \left(A_j^*(\phi) + B_j^*(\phi)\nu_j \right) + i\phi(r-q)\tau \right] \right) d\phi \quad (4.2.1)$$

With $A_j(\phi) = A_j^*(-\phi)$ and $B_j(\phi) = B_j^*(-\phi)$, $p(s)$ can be rewritten as

$$p(s) = \frac{1}{\pi} \int_0^{\infty} \mathrm{Re}\left(e^{-i\phi s} e^{-i\phi(r-q)\tau} G(\phi) \right) d\phi \quad (4.2.2)$$

So the log-moneyness probability density function $p(s)$ is just the Fourier Transform of the fundamental transform $e^{-i\phi(r-q)\tau} G(\phi)$ in (4.1.14). From this formula, it is easily seen that the log-moneyness probability density function $p(s)$ depends on $\sum_{j=1}^{N} B_j(\phi)\nu_j$, not on $\sum_{j=1}^{N} \nu_j$.

We will use a Double-Heston model example to illustrate how two variance processes influence the log-moneyness probability density function $p(s)$. Assume $\kappa_1 = 2.13$, $\kappa_2 = 5.49$, $\theta_1 = 0.085$, $\theta_2 = 0.058$, $\xi_1 = 0.27$, $\xi_2 = 0.1$, $\nu_1(0) = 0.091$, $\nu_2(0) = 0.065$, $\rho_1 = -0.86$, $\rho_2 = 0.96$, $r = 0.07$, $q = 0$, and $T = 1$ as the original parameters. Figure (4.2.1) gives some examples of $p(s)$ under the influence of different parameters. The parameters

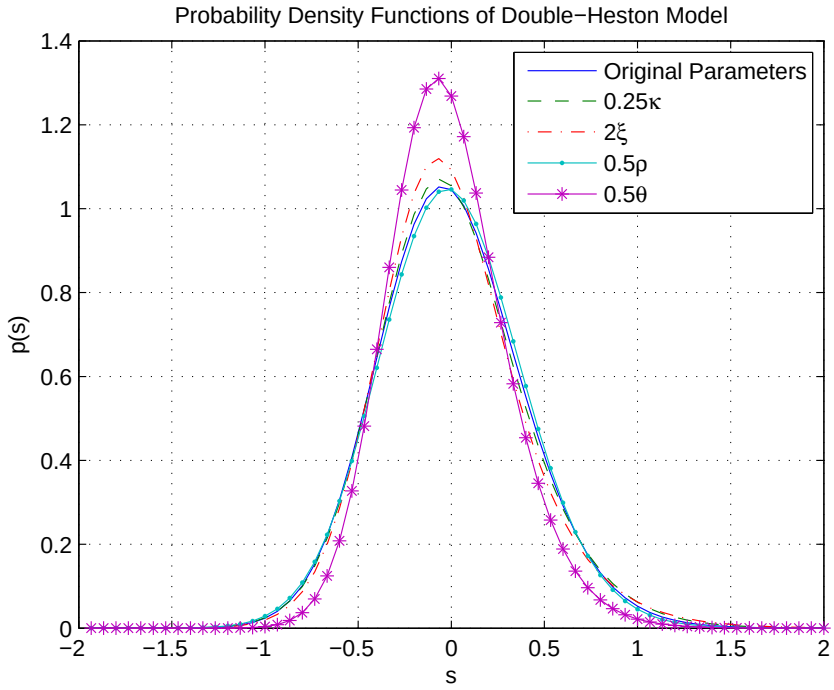

FIGURE 4.2.1. Examples of $p(s)$ Changing with Parameter Changes. (See Text for Parameter Values)

are assumed unchanged in each curve in this figure except the one being indicated. For example, 0.25κ means that both κ_1 and κ_2 are reduced to a quarter of their original values.

Kainth et al [54] applied the Double-Heston model to the FX European and DNT (a barrier-type product called "Double No Touch") products and found the model captures the smile feature of the market and it can reproduce the prices very well. Kainth et al [54] argued that the Multi-Heston model echoes the market that appears to display more than one volatility processes in its dynamics. It can accommodate more complex mean reverting properties. For example, the volatility may revert on multiple time scales, one fast and one slow. The stochastic vol may quickly revert to one long vol which itself is slowly reverting [5]. The inclusion of more mean-reverting variance processes could potentially help to describe the underlying's stochastic volatility movement in more detail. Christoffersen et al [20] demonstrated that the double-Heston model is more flexible in modelling the level and slope of the smile that empirically known fluctuate largely independently with each other, and it is more flexible in modelling the volatility term structure. Through empirical studies, Christoffersen et al [20] found that in the single factor Heston model, "the conditional skewness and kurtosis are very strongly linked to the trend and variation in the conditional variance" while such a relationship is much weaker in the double-Heston

model, hence benefitting the modelling of the largely independently fluctuating level and slope of the smile. Christoffersen et al [20] also pointed out that the resulting differences between the one-factor and two factor models strongly depend on what parameter estimates are used. Indeed, from (4.2.2), we can see that the parameters of the Multi-Heston model affect the probability density distribution function is a rather complex fashion through the integral of the nonlinear function G. More empirical studies may reveal more applications where the Multi-Heston model can have an advantage.

4.3. Numerical Implementation with the Kahl-Jäckel Algorithm

Now let us look at how to implement the semi-analytical solution of the Multi-Heston model. To this end, we will extend the Kahl-Jäckel algorithm [53], which uses a "rotation count" algorithm to handle the complex discontinuity which originates from the component $\ln\left(\frac{c^* e^{d^*\tau}-1}{c^*-1}\right)$ in (4.1.6).

The first type of error may arise when taking the logarithm of a complex number produces multiple values. Let $z = re^{i(t+2\pi n)}$ be a complex number and $t \in (-\pi, \pi]$ be the phase in the principal branch. The logarithm of z is $\ln|r| + i\,(t + 2\pi n)$. If we restrict the phase to the principal branch or any single branch, a discontinuity will occur. The second type of discontinuity is caused by the branch switching of the complex power function. For example, we know $e^{i\pi} = e^{-i\pi}$ but $i = \left(e^{i\pi}\right)^{1/2} \neq \left(e^{-i\pi}\right)^{1/2} = -i$. If a 2π is added to the phase of $e^{-i\pi}$, we obtain $i = \left(e^{i\pi}\right)^{1/2} = \left(e^{-i\pi+i2\pi}\right)^{1/2} = i$ while keeping $e^{i\pi} = e^{-i\pi} = e^{-i\pi+i2\pi}$. This example illustrates how important it is to keep an eye on the phase while doing complex operations.

Kahl-Jäckel [53] presented an algorithm called "rotation count" as follows:

$$V_{call} = e^{-r\tau}\int_0^1 Y(y)dy \tag{4.3.1}$$

where

$$\begin{cases} Y(y) &= \dfrac{1}{2}(F-K) + \dfrac{Ff_1\left(-\frac{\ln y}{\gamma}\right) - Kf_2\left(-\frac{\ln x}{\gamma}\right)}{y\pi\gamma} \\ f_1 &= \mathrm{Re}\left(\dfrac{e^{i\frac{\ln y}{\gamma}\ln K}\varphi(-\frac{\ln y}{\gamma}-i)}{i\left(-\frac{\ln y}{\gamma}\right)F}\right) \\ f_2 &= \mathrm{Re}\left(\dfrac{e^{i\frac{\ln y}{\gamma}\ln K}\varphi(-\frac{\ln y}{\gamma})}{i\left(-\frac{\ln y}{\gamma}\right)}\right) \end{cases} \tag{4.3.2}$$

and

$$\gamma = \frac{\sqrt{1-\rho^2}}{\xi}\left(\nu + \kappa\theta\tau\right) \tag{4.3.3}$$

This formula is identical to (4.1.4). Kahl-Jäckel made a coordinate transform $y = e^{-\phi\gamma}$, where $\gamma > 0$. Thus y has the range from 0 to 1 while $\phi \in [0, \infty]$.

According to the Kahl-Jäckel algorithm, the function of φ is obtained by

$$\begin{cases} d & = & \sqrt{(\rho\xi\phi i - \kappa)^2 + \xi^2(\phi i + \phi^2)} \\ c & = & \dfrac{\kappa - \rho\xi\phi i + d}{\kappa - \rho\xi\phi i - d} \\ t_c & = & \arg(c) \\ G_D & = & c - 1 \\ m & = & \mathrm{int}[\dfrac{t_c + \pi}{2\pi}] \\ G_N & = & ce^{d\tau} - 1 \\ n & = & \mathrm{int}[\dfrac{t_c + \mathrm{Im}(d)\tau + \pi}{2\pi}] \\ \ln G & = & \ln(\mathrm{abs}(G_N)/\mathrm{abs}(G_D)) + i(\arg(G_N) - \arg(G_D) + 2\pi(n-m)) \\ B^* & = & \dfrac{\kappa - \rho\xi\phi i + d}{\xi^2}\left(\dfrac{e^{d\tau} - 1}{ce^{d\tau} - 1}\right) \\ A^* & = & \dfrac{\kappa\theta}{\xi^2}\left[(\kappa - \rho\xi\phi i + d)\,\tau - 2\ln G\right] \\ \varphi & = & \exp\left(A^* + B^*\nu + i\phi\ln(F)\right) \end{cases} \tag{4.3.4}$$

where int [ ] means Gauss's integer bracket (or the floor function).

It has been previously proved that the characteristic function of the Multi-Heston model can be synthesized by A_j^*, B_j^* as if they are obtained with the Heston model formula. So

we have

$$\begin{cases} d_j & = & \sqrt{(\rho_j\xi_j\phi i-\kappa_j)^2+\xi_j^2(\phi i+\phi^2)} \\ c_j & = & \dfrac{\kappa_j-\rho_j\xi_j\phi i+d_j}{\kappa_j-\rho_j\xi_j\phi i-d_j} \\ t_c^j & = & \arg(c_j) \\ G_D^j & = & c_j-1 \\ m_j & = & \text{int}[\dfrac{t_c^j+\pi}{2\pi}] \\ G_N^j & = & c_j e^{d_j\tau}-1 \\ n_j & = & \text{int}[\dfrac{t_c^j+\text{Im}(d_j)\tau+\pi}{2\pi}] \\ \ln G_j & = & \ln(\text{abs}(G_N^j)/\text{abs}(G_D^J))+i(\arg(G_N^j)-\arg(G_D^j)+2\pi(n_j-m_j)) \\ B_j^* & = & \dfrac{\kappa_j-\rho_j\xi_j\phi i+d_j}{\xi_j^2}\left(\dfrac{e^{d_j\tau}-1}{c_j e^{d_j\tau}-1}\right) \\ A_j^* & = & \dfrac{\kappa_j\theta_j}{\xi_j^2}\left[(\kappa_j-\rho_j\xi_j\phi i+d_j)\,\tau-2\ln G_j\right] \\ \varphi & = & \exp\left(\sum_{j=1}^{N}(A_j^*+B_j^*\nu_j)+i\phi\ln(F)\right) \end{cases} \tag{4.3.5}$$

Although any positive function γ seems to be fine with (4.3.2), Kahl-Jäckel choose the form of (4.3.3) because it best reflects the asymptotic decay of the integrand in (4.3.1). The function $f(\phi)$ has the following asymptotic behaviour as ϕ tends to infinity:

$$\lim_{\phi\to\infty} f(\phi) \propto e^{-\phi\text{Re}\left[\lim_{\phi\to\infty}\frac{(A^*+B^*\nu)}{\phi}\right]}\phi^{-1} \tag{4.3.6}$$

Kahl-Jäckel selected $\gamma = \text{Re}\left[\lim_{\phi\to\infty}\frac{(A^*+B^*\nu)}{\phi}\right]$ to reflect the fact that asymptotically $f(\phi)$ decays at least exponentially.

Similarly, with the Multi-Heston model, it can be proved with the Kahl-Jäckel approach that

$$\lim_{\phi\to\infty} f(\phi) \propto e^{-\phi Re\left[\lim_{\phi\to\infty}\frac{\sum_{j=1}^{N}\left(A_j^*+B_j^*\nu_j\right)}{\phi}\right]}\phi^{-1} \tag{4.3.7}$$

Therefore, we choose

$$\gamma_{MultiHeston} \equiv Re\left[\lim_{\phi\to\infty}\frac{\sum_{j=1}^{N}\left(A_j^*+B_j^*\nu_j\right)}{\phi}\right] = \sum_{j=1}^{N}\left(\frac{\sqrt{1-\rho_j^2}}{\xi_j}\left(\nu_j+\kappa_j\theta_j\tau\right)\right) \tag{4.3.8}$$

The important part of the Kahl-Jäckel approach is to get the characteristic function φ through the "rotation counting" algorithm. To price an option (given the condition that

the payoff function does not depend on volatility), we can just get φ and then use (4.1.21) by plugging in the Fourier Transform of the payoff.

The generalised Multi-Heston formula (4.1.21) can be re-arranged into

$$\begin{aligned} V &= \frac{e^{-r\tau}}{2\pi}\int_{-\infty}^{+\infty}\tilde{W}\left(i\phi_I-\phi,\,0\right)\varphi(-i\phi_I+\phi)d\phi \\ &= e^{-r\tau}\int_{0}^{+\infty}f\left(\phi\right)d\phi \end{aligned}$$

where

$$f\left(\phi\right)=\frac{1}{2\pi}\left[\tilde{W}\left(i\phi_I-\phi,\,0\right)\varphi(-i\phi_I+\phi)+\tilde{W}\left(i\phi_I+\phi,\,0\right)\varphi(-i\phi_I-\phi)\right]. \tag{4.3.9}$$

Through the same coordinate transform $\phi=-\dfrac{\ln y}{\gamma}$ as in the method of Kahl-Jäckel [53], we obtain

$$V=e^{-r\tau}\int_0^1\frac{f(-\frac{\ln y}{\gamma})}{y\gamma}dy \tag{4.3.10}$$

A numerical integration over the range of [0, 1] can be implemented very easily. This range seems to be better than the range of $[-\infty,\,\infty]$ because no guess about the cutoff is necessary with the [0, 1] range. On Matlab, the numerical integration over this range can be implemented with extremely fast speed. Using the Kahl-Jäckel [53] approach, we have verified that the prices obtained through the formulas (4.1.4) and (4.1.21) are exactly the same. (In the numerical tests, the results are to the accuracy 10^{-10} set for the Matlab function quadl()). Next, we will give the comparison between the analytical solution and the Monte-Carlo solution of the Multi-Heston model.

N	κ_j	θ_j	ξ_j	$\nu_j\left(0\right)$	ρ_j	$\frac{\kappa_j\theta_j}{\xi_j^2}$
1	2.13	0.085	0.27	0.091	-0.86	2.48
2	5.49	0.058	0.1	0.065	0.96	31.84
3	1.67	0.038	0.2	0.055	-0.76	1.59
4	1.58	0.028	0.1	0.045	0.91	4.42
5	1.67	0.018	0.2	0.075	0.22	0.75
6	1.13	0.075	0.17	0.036	-0.83	2.93
7	4.31	0.047	0.15	0.045	0.2	9.00
8	3.89	0.088	0.13	0.012	-0.19	20.26
9	2.91	0.052	0.18	0.045	0.55	4.67
10	4.77	0.018	0.19	0.03	0.15	2.38

TABLE 4.2. Parameters for the Multi-Heston model test

N	**Anal.**	**C.I.**(10^4)	**C.I.**(10^5)	**A.E.**(10^4)	**A.E.**(10^5)	**R.E.**(10^4)	**R.E.**(10^5)
1	0.3493	0.3575±0.0208	0.3464±0.0065	0.0082	0.0029	1.1800	1.3568
2	0.5459	0.5397±0.0354	0.5385±0.0110	0.0062	0.0074	0.5256	2.0127
3	0.6603	0.6743±0.0431	0.6604±0.0135	0.0140	0.0001	0.9785	0.0166
4	0.7472	0.7324±0.0502	0.7454±0.0157	0.0148	0.0019	0.8841	0.3534
5	0.8442	0.8763±0.0616	0.8428±0.0184	0.0321	0.0014	1.5621	0.2244
6	0.9388	0.9251±0.0643	0.9460±0.0208	0.0137	0.0072	0.6377	1.0414
7	1.0212	1.0314±0.0725	1.0311±0.0232	0.0101	0.0098	0.4198	1.2698
8	1.1324	1.1529±0.0844	1.1392±0.0264	0.0205	0.0067	0.7286	0.7633
9	1.2080	1.1754±0.0890	1.2148±0.0291	0.0327	0.0068	1.1010	0.6978
10	1.2375	1.1410±0.0892	1.2258±0.0300	0.0965	0.0117	3.2468	1.1756

TABLE 4.3. Error behaviour as the number of Monte-Carlo simulation samples increases from 10^4 to 10^5

N	**Anal.**	**C.I.**(10^{-1})	**C.I.**(10^{-2})	**A.E.**(10^{-1})	**A.E.**(10^{-2})	**R.E.**(10^{-1})	**R.E.**(10^{-2})
1	0.3493	0.3429±0.0063	0.3464±0.0065	0.0064	0.0029	3.0551	1.3568
2	0.5459	0.5383±0.0107	0.5385±0.0110	0.0076	0.0074	2.1242	2.0127
3	0.6603	0.6539±0.0130	0.6604±0.0135	0.0064	0.0001	1.4732	0.0166
4	0.7472	0.7366±0.0151	0.7454±0.0157	0.0106	0.0019	2.1185	0.3534
5	0.8442	0.8410±0.0176	0.8428±0.0184	0.0032	0.0014	0.5445	0.2244
6	0.9388	0.9332±0.0199	0.9460±0.0208	0.0056	0.0072	0.8505	1.0414
7	1.0212	1.0195±0.0222	1.0311±0.0232	0.0017	0.0098	0.2284	1.2698
8	1.1324	1.1483±0.0251	1.1392±0.0264	0.0159	0.0067	1.8939	0.7633
9	1.2080	1.2366±0.0278	1.2148±0.0291	0.0286	0.0068	3.0847	0.6978
10	1.2375	1.2511±0.0283	1.2258±0.0300	0.0135	0.0117	1.4308	1.1756

TABLE 4.4. Error behaviour as δt is refined from 0.1 to 0.01

Table (4.3) is an example comparing the analytical result (denoted by "Anal.") and Monte-Carlo confidence interval (denoted by "C.I.") for a vanilla call option based on the N dimensional Multi-Heston model. The C.I. (m) columns are the confidence intervals on 3 standard deviations with m Monte-Carlo simulation samples, which are expressed by Mean $\pm\, 3\dfrac{\text{St.Dev.}}{\sqrt{m}}$. The A.E.$(m)$ columns give the error results with m Monte-Carlo simulation samples, which are expressed by abs(Anal.-Mean) the absolute error between the analytical solution and the Monte-Carlo mean whereas the R.E.(m) columns give the relative errors defined by A.E./$(\dfrac{\text{St.Dev.}}{\sqrt{m}})$. The Multi-Heston analytical and Monte-Carlo results corresponding to a specific N value are computed with the parameters listed in Table (4.2) for dimensions from the first to the Nth included. The parameters are chosen to satisfy $\dfrac{\kappa_j\theta_j}{\xi_j^2} > \dfrac{1}{2}$ (see [72, 54]) so that the ν_j will not go to zero. The other parameters

used in this test are $r = 0.07$, $q = 0$, $K = 6.15$, $S = 5.15$, and $T = 1$. The Euler Monte-Carlo scheme is used with the time interval $\delta t = 0.01$. We follow Glasserman [38] for the Monte-Carlo algorithm.

Table (4.4) shows the error comparison between $\delta t = 0.1$ and $\delta t = 0.01$ for 10^5 Monte-Carlo simulation samples. Shaw [89] suggested that the Milstein scheme (see [38]) if used with the Heston model will have the difficulty of computing the "Levy Stochastic Area" for the first Heston SDE involving dS_t because the stochastic variance ν changes within one time-step evolution of S_t. Although pointing to an $O\left((\delta t)^{3/2}\right)$ scheme proposed by Kloeden and Platen [58], Shaw [89] recommended that the Euler method be used with finer δt as the Euler method is much easy to implement. Table (4.4) shows that the discretisation error resulting from a coarser $\delta t = 0.1$ is obvious; When the δt is fine enough as in the case of $\delta t = 0.01$, the relative errors look very much in the regular range and give the indication that $\delta t = 0.01$ is a sensible choice for the Euler simulation.

4.3.1. Easy Computation of Greeks and a Candidate for Karasinski. Shaw [89] suggested with taking Greeks with respect to spot using the Heston formula is very easy. It can be seen in (4.1.20) that we only need to multiply the integrand by $-\frac{i\phi}{S}$ and $-\frac{\phi^2}{S^2}$ to get delta and gamma, respectively. Gatheral [36] suggested that the Greeks with respect to the variance ν and spot S are very easy to compute with the Heston formula because A^* and B^* in the characteristic function (4.3.4) contain neither variance nor spot. The same statement can be applied to the Multi-Heston formula as well. The rotation counting of the Kahl-Jäckel algorithm will only need to run once for A_j^* and B_j^* in (4.3.5).

The first and second order sensitivities of φ in (4.3.5) with respect to ν_j and S are given

$$\begin{aligned}
&\frac{\partial \varphi}{\partial S} = \frac{i\phi}{S}\varphi,\ \frac{\partial \varphi}{\partial \nu_j} = B_j^*\varphi \\
&\frac{\partial^2 \varphi}{\partial S^2} = -\frac{\phi^2 + i\phi}{S^2}\varphi,\ \frac{\partial^2 \varphi}{\partial \nu_j^2} = \left(B_j^*\right)^2\varphi \\
&\frac{\partial^2 \varphi}{\partial S \partial \nu_j} = \frac{i\phi B_j^*}{S}\varphi
\end{aligned} \tag{4.3.11}$$

With (4.3.11), the explicit formulas for the Greeks of V can be easily written but since the combined expressions of them are lengthy and do not offer new insight, we choose not to include them here.

A downside of the Heston-family stochastic models is slow to calibrate and difficult to reach high accuracy in calibration. However, because both price and Greeks can be computed with the analytical formulas with constant model parameters, the Heston-family stochastic models with constant parameters can be quickly computed with these formulas. The characteristic of the constant-parameter Heston-family models being difficult-to-calibrate but fast-to-compute will be echoed in the later chapter about the Karasinski method, which could help improve accuracy in these models.

CHAPTER 5

The Vanna-Volga Method

5.1. Some Background on FX Conventions

An exchange rate S means the amount of domestic currency needed to buy one unit of the foreign currency [104],

$$S = \frac{\text{Domestic Currency}}{\text{Foreign Currency}} \tag{5.1.1}$$

In the Black-Scholes model, an exchange rate S is modelled by the geometric Brownian motion [104]

$$dS = (r_d - r_f)\, Sdt + \sigma S dW \tag{5.1.2}$$

where r_d is the domestic interest rate, r_f the foreign interest rate, σ the Black-Scholes implied volatility, and W a Wiener process.

The difference between the domestic currency and the foreign currency has nothing to do with a trader's nation. It is used only to denote the domestic currency as the numeraire currency. For example, in the quote of USD-JPY, yen is the domestic currency and dollar the foreign currency; the option price computed by the Black-Scholes formula is given in yen per unit of dollar (termed d pips). In FX trading, however, an option price can be given by 6 different ways, namely domestic cash (d), foreign cash (f), % domestic (%d), % foreign (%f), domestic pips (d pips) and foreign pips (f pips). Their relationships are summarised below [104],

$$\begin{aligned} &\text{d pips} = S \times \%\text{f} = K \times \%\text{d} = KS \times \text{f pips} \\ &\text{d} = \text{Notional} \times \%\text{d} \\ &\text{f} = \text{Notional} \times \%\text{f} \end{aligned} \tag{5.1.3}$$

where K is strike. The Black-Scholes price yields a value expressed in d pips.

A vanilla option, whose price will be denoted by U in this thesis from now on, has the payoff of $(S_T - K)^+$ if it is a call option or $(K - S_T)^+$ if it is a put. At some strikes, the vanilla options have observable market quotes. In the FX market, the liquid vanilla options

are quoted by their Black-Scholes implied volatilities but these volatilities are not quoted directly by strikes; rather, the volatilities are quoted using the Delta values computed by the Black-Scholes formula

$$\Delta = \phi e^{-r_f \tau} \mathcal{N}\left(\phi d_+\right), \tag{5.1.4}$$

where $\mathcal{N}$ is the normal cumulative function,

$$d_+ = \frac{\ln\left(S/K\right) + \left(r_d - r_f + \sigma^2/2\right)\tau}{\sigma\sqrt{\tau}},$$

K is the strike, σ is the Black-Scholes implied vol, τ is the time to maturity , and

$$\begin{cases} \phi = 1, & \text{Call,} \\ \phi = -1, & \text{Put.} \end{cases}$$

There are just a few such liquid volatilities, namely 10Δ put (a put option with the Black-Scholes Delta being -10%), 25Δ put (a put option with the Black-Scholes Delta being -25%), at-the-money (ATM) (the strike where $\Delta_{call} = -\Delta_{put}$), 25Δ call (a call option with the Black-Scholes Delta being 25%), and 10Δ call (a call option with the Black-Scholes Delta being 10%). In FX, the standard notion of ATM [104] in the market quotes is the Delta parity where $\Delta_{call} = -\Delta_{put}$ holds on K_{ATM} although the value parity notion of ATM (where call and put option values are equal on K_{ATM}) also exists. The equity option notion of ATM means $S = K$. In this thesis, we will follow the market standard notion of ATM. Given the Delta parity, the ATM strike [82] is

$$K_{ATM} = S \exp\left(\left(r_d - r_f + \frac{\sigma_{ATM}^2}{2}\right)\tau\right) \tag{5.1.5}$$

where S is the current spot, σ_{ATM} is the ATM Black-Scholes implied vol, and τ is the time to maturity.

Among the five liquid strikes, the three strikes namely 25Δ put, ATM, 25Δ call are most liquid. The σ–Δ pairs are just a mapping for the σ–K pairs for convenience; In essence, the σ–Δ pairs are just a way of expressing the σ–K pairs. Using the σ–Δ pairs help to standardise the vanilla products across different currency markets. Compared with non-intuitive strikes of vastly different scales in different currency markets, quoting the same set of liquid Δ's is obviously a better language.

Although the strikes of vanillas are not explicitly given by the market quotes, the strike of a vanilla can be easily inferred from the Delta and the market Black-Scholes implied

volatility associated with it using [104]

$$K = S \exp\left[-\phi \mathcal{N}^{-1}\left(\phi \Delta e^{r_f \tau}\right) \sigma \sqrt{\tau} + \left(r_d - r_f + \frac{\sigma^2}{2}\right) \tau\right]. \tag{5.1.6}$$

This formula is simply inverted from the formula of the Black Scholes Delta (5.1.4). The discrete liquid strikes are often referred to as the *vol pillars*.

5.1.0.1. *RR and BF - Simple Skew and Kurtosis Indicators.* The skew and kurtosis are statistically defined as the third central and fourth central moments respectively [78]. The normalised skew (denoted by μ_3) and kurtosis (denoted by μ_4) of a random variable X with standard deviation σ_X are defined by

$$\begin{aligned} \mu_3 &= \frac{E\left[\left(X - E\left(X\right)\right)^3\right]}{\sigma_X^3}, \\ \mu_4 &= \frac{E\left[\left(X - E\left(X\right)\right)^4\right]}{\sigma_X^4}. \end{aligned}$$

When X has a normal probability density, $\mu_4 = 3$ holds, so $\mu_4 - 3$ is called the excess kurtosis. To use the statistical definitions directly on the state price $X = S_T$ of the underlying will not be easy in practice. Since there are only a handful of such liquid pillars, it is convenient to define some parameters based on the liquid vol pillars to describe the skew and kurtosis of the underlying return as well as the slope and curvature of the implied volatility smile. We define

$$\begin{aligned} RR &= \sigma_c - \sigma_p \\ BF &= \frac{\sigma_c - 2\sigma_{ATM} + \sigma_p}{2} \end{aligned} \tag{5.1.7}$$

where σ_c is the implied volatility of an out-of-the-money call, σ_p the implied volatility of an out-of-the-money put, and σ_{ATM} the implied volatility of the at-the-money option. RR is called the risk reversal and BF the butterfly.

Usually the volatility of 25Δ call is used for σ_c and the volatility of 25Δ put for σ_p.

From (5.1.7), $\sigma_{25\Delta Call}$ and $\sigma_{25\Delta Put}$ are given by

$$\begin{aligned} \sigma_{25\Delta Call} &= \sigma_{ATM} + BF + 0.5RR \\ \sigma_{25\Delta Put} &= \sigma_{ATM} + BF - 0.5RR \end{aligned} \tag{5.1.8}$$

Sometimes we run into situations when risk reversals and butterflies are expressed with vanilla prices rather than volatilities. For the sake of consistency, in this thesis, RR and BF are reserved for risk reversals and butterflies expressed by volatilities whereas $\widetilde{RR}$ and

$\widetilde{BF}$ are used to denote risk reversals and butterflies expressed not by volatilities but by prices:

$$\begin{aligned} \widetilde{RR} &= U_{25\Delta Call} - U_{25\Delta Put} \\ \widetilde{BF} &= \frac{U_{25\Delta Call} - 2U_{ATM} + U_{25\Delta Put}}{2} \end{aligned} \tag{5.1.9}$$

Carr and Wu [14] and Wu [102] state that in the quotes of FX options RR is used to describe the skew μ_3 and BF indicates the level of the kurtosis μ_4. This can be easily seen from the quadratic implied vol fitting function given by Wu [102]. Wu [102] suggested that the Black-Scholes implied vol smile can be approximately fitted by the following quadratic function using the skew and kurtosis of the underlying return as its coefficients:

$$\sigma(k) \approx \sigma_{ATM}\left(1 + \frac{\mu_3}{6}k + \frac{\mu_4 - 3}{24}k^2\right) \tag{5.1.10}$$

where

$$k = \frac{\ln(K/F)}{\sigma_{ATM}\sqrt{\tau}},\ K:\ \text{Strike},\ F:\ \text{Forward Price}$$

Let the three liquid log-strikes be denoted by k_p, k_{ATM}, and k_c, which correspond to the implied vols σ_p, σ_{ATM}, and σ_c, respectively. With the quadratic function (5.1.10), RR and BF can be approximately written as

$$\begin{aligned} RR &\approx \left[\frac{\mu_3}{6}(k_c - k_p) + \frac{\mu_4 - 3}{24}\left(k_c^2 - k_p^2\right)\right]\sigma_{ATM} \\ BF &\approx \left[\frac{\mu_4 - 3}{24}\left(\frac{\left(k_c^2 + k_p^2\right)}{2} - k_{ATM}^2\right) + \frac{\mu_3}{6}\left(\frac{(k_c + k_p)}{2} - k_{ATM}\right)\right]\sigma_{ATM} \end{aligned} \tag{5.1.11}$$

In the RR equality, the $\frac{\mu_4 - 3}{24}\left(k_c^2 - k_p^2\right)$ term is usually far smaller than the $\frac{\mu_3}{6}(k_c - k_p)$ term. In the BF equality, k_{ATM} can be approximately viewed as the middle point of the two strikes k_c and k_p on its two sides, hence $(k_c + k_p)/2 - k_{ATM} \approx 0$. So the equalities in (5.1.11) can be simplified:

$$\begin{aligned} RR &\approx \frac{\mu_3}{6}(k_c - k_p)\sigma_{ATM} \\ BF &\approx \frac{\mu_4 - 3}{24}\left(\frac{\left(k_c^2 + k_p^2\right)}{2} - k_{ATM}^2\right)\sigma_{ATM} \end{aligned} \tag{5.1.12}$$

The approximate linear relations between RR and μ_3 and between BF and μ_4 suggest that it is a sensible approximation to use RR and BF as simple measures for the skew and kurtosis of the underlying return, respectively.

5.2. Vega Hedging and the Vanna-Volga Method

First, we will introduce some Greeks that are relevant in the discussion in this chapter. The Black-Scholes Greeks Vega, Volga, and Vanna are defined by

$$\begin{aligned} \text{Vega} &= \frac{\partial V^{BS}}{\partial \sigma} \\ \text{Volga} &= \frac{\partial^2 V^{BS}}{\partial \sigma^2} \\ \text{Vanna} &= \frac{\partial^2 V^{BS}}{\partial \sigma \partial S}. \end{aligned} \tag{5.2.1}$$

where V^{BS} is the Black-Scholes option price and σ is the Black-Scholes implied volatility.

5.2.1. Hedging Against Volatility Risk. FX option products are built on exchange rates, which are fast moving and inclined to have significant skews. The flat vol Black-Scholes model is obviously not suitable for the task of accurate estimation of the mark-to-market price. Stochastic or local vol models can be used but high quality calibration is often required for these models. There are two problems: firstly, calibration takes time to complete and the market may drift a bit since the last calibration or even during calibration; secondly, the option prices can be very sensitive to spot and volatility changes, and even small errors in estimates of spot and vol may result in unacceptable errors in prices.

A common industrial practice that reduces the risks of a model is through hedging. Hedging the risk of spot is straightforward because the spot S (the currencies in the FX market) is a tradable asset. As volatility σ is not directly traded, the difficult part is the hedging of the volatility risk. In the FX market, in order to hedge the spot and volatility risks of a target option whose price is denoted by V, one can construct a portfolio Π^{mkt} including some currencies represented by the exchange rate S and in the simple case a number of just one vanilla option whose price is denoted by U^{mkt}.

This portfolio can be written as

$$\Pi^{mkt} = V^{mkt} - \Delta S - \omega U^{mkt} \tag{5.2.2}$$

where Δ and ω are the numbers of currencies and the vanilla option held in the portfolio, respectively; mkt denotes "market". Note U^{mkt} is observable in the market and V^{mkt} is not.

The relationship between the change of value of Π^{mkt} and the changes in the values of S and U^{mkt} is

$$d\Pi^{mkt} = dV^{mkt} - \Delta dS - \omega dU^{mkt} \tag{5.2.3}$$

The exact dynamics of the market are unknown. Therefore Δ and ω have to be computed with a specific model that is hoped to approximate reality. Assume both S and its instantaneous vol follow diffusion processes

$$\begin{aligned} dS &= (...)\, dt + \sigma dW_1 \\ d\sigma &= (...)\, dt + (...)\, dW_2 \end{aligned}$$

where $dW_1 dW_2 = \rho dt$ and $(...)$ means the coefficient is not specified. Assume V^{mod} and U^{mod} are the option prices under this model. Assume also V^{mod} and U^{mod} are twice differentiable w.r.t. S and σ.

We rearrange (5.2.2) into

$$\Pi^{mkt} = V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) + \Pi^{mod}$$

where $\Pi^{mod} = V^{mod} - \Delta S - \omega U^{mod}$.

Ignoring the second-order terms, $d\Pi$ can be expanded through Itô's Lemma and will have the following stochastic terms

$$\begin{aligned} d\Pi^{mkt}\big|_{\text{Stoch.}} = \; & d\left[V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) \right] + \\ & \left(\frac{\partial V^{mod}}{\partial S} - \Delta - \omega \frac{\partial U^{mod}}{\partial S} \right) dS + \left(\frac{\partial V^{mod}}{\partial \sigma} - \omega \frac{\partial U^{mod}}{\partial \sigma} \right) d\sigma \end{aligned} \tag{5.2.4}$$

When Δ and ω are chosen to satisfy

$$\begin{aligned} \Delta &= \frac{\partial V^{mod}}{\partial S} - \omega \frac{\partial U^{mod}}{\partial S} \\ \omega &= \frac{\partial V^{mod}}{\partial \sigma} \Big/ \frac{\partial U^{mod}}{\partial \sigma} \end{aligned} \tag{5.2.5}$$

the dS and $d\sigma$ terms in $d\Pi^{mod}$ will be eliminated and $d\Pi^{mkt}\big|_{\text{Stoch.}}$ will follow

$$d\Pi^{mkt}\Big|_{\text{Stoch.}} \approx d\left[V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) \right] \tag{5.2.6}$$

To have a perfect hedge, $d\Pi^{mkt}\big|_{\text{Stoch.}}$ needs to be zero as well. That will require

$$d\left[V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) \right] = 0 \tag{5.2.7}$$

There is no guarantee that $\omega = \dfrac{\partial V^{mod}}{\partial \sigma} / \dfrac{\partial U^{mod}}{\partial \sigma}$ makes (5.2.7) true. So a perfect hedge is impossible unless the market and the model coincide. However, if V^{mod} and U^{mod} are computed with a model closely following the market but with a slightly wrong σ, the expression (5.2.5) can be written as

$$\omega = \frac{\partial V^{mod}}{\partial \sigma} / \frac{\partial U^{mod}}{\partial \sigma} \approx \left(\frac{V^{mkt} - V^{mod}}{\sigma^{mkt} - \sigma^{mod}} \right) / \left(\frac{U^{mkt} - U^{mod}}{\sigma^{mkt} - \sigma^{mod}} \right),$$

hence

$$V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) \approx 0. \tag{5.2.8}$$

and therefore

$$d\Pi^{mkt}\Big|_{\text{Stoch.}} \approx d \left[V^{mkt} - V^{mod} - \omega \left(U^{mkt} - U^{mod} \right) \right] \approx 0.$$

When the stochastic terms are eliminated, in the ideal case described above, $d\Pi^{mkt}$ will grow at the risk-free rate r,

$$d\Pi^{mkt} \approx r\Pi^{mkt} dt \tag{5.2.9}$$

For real cases, however, the hedged portfolio will not be entirely free of risk.

The equality (5.2.8) means that $V^{mod} + \omega \left(U^{mkt} - U^{mod} \right)$ can be used to approximate the market price. Hence a good candidate of mark-to-market price is

$$\widehat{V}^{mkt} = V^{mod} + \omega \left(U^{mkt} - U^{mod} \right) \tag{5.2.10}$$

where ω is given by (5.2.5).

When the model is the Black-Scholes model, and σ is the ATM Black-Scholes implied vol, (5.2.10) is rewritten as

$$\widehat{V}^{mkt} = V^{BS}\big|_{\sigma_{ATM}} + \omega \left(U^{mkt} - U^{BS}\big|_{\sigma_{ATM}} \right) \tag{5.2.11}$$

where BS denotes the Black-Scholes model and

$$\omega = \frac{\partial V^{BS}}{\partial \sigma}\bigg|_{\sigma_{ATM}} / \frac{\partial U^{BS}}{\partial \sigma}\bigg|_{\sigma_{ATM}} \tag{5.2.12}$$

This is just Vega hedging under the assumption that the model is closely following the market.

If (5.2.11) is to be used to price an option in the FX market, a practical requirement is that this pricing method must be able to produce the market prices on the three of the most liquid strikes described earlier, namely 25Δ put, ATM, 25Δ call. With only one market input variable U^{mkt}, (5.2.11) obviously cannot fit the prices of all three liquid strikes. Naturally, (5.2.11) can be extended to

$$\widehat{V}^{mkt} = V^{BS}\big|_{\sigma_{ATM}} + \sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \tag{5.2.13}$$

With three market input variables U_k^{mkt}, $1 \leqslant k \leqslant 3$, it is possible that the mark-to-market price V will be able to fit the prices of all three liquid strikes. So the remaining issue is how to determine the weights ω_k. With one market input, ω in (5.2.12) is computed by making $V^{BS}\big|_{\sigma_{ATM}} - \omega\, U^{BS}\big|_{\sigma_{ATM}}$ Vega neutral. With three market input variables as in (5.2.13), in relation with the one market input case, a way of computing the weights ω_k is by making $V^{BS}\big|_{\sigma_{ATM}} - \sum_{k=1}^{3} \omega\, U^{BS}\big|_{\sigma_{ATM}}$ Vega, Volga, and Vanna neutral. The formula (5.2.13) combined with ω_k computed by making $V^{BS}\big|_{\sigma_{ATM}} - \sum_{k=1}^{3} \omega\, U^{BS}\big|_{\sigma_{ATM}}$ Vega, Volga, and Vanna neutral is called the Vanna-Volga (VV) method, which will be discussed next.

5.2.2. Castagna and Mercurio Representation of the Vanna-Volga Method. The above-mentioned Vanna-Volga (VV) method was first mentioned by Lipton and McGhee [66]. The option price computed by the VV method is

$$V^{VV} = V^{BS}\big|_{\sigma_{ATM}} + \sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \tag{5.2.14}$$

where the weights ω_k satisfy

$$\begin{aligned}
\left.\frac{\partial V^{BS}}{\partial \sigma}\right|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \left.\frac{\partial U_k^{BS}}{\partial \sigma}\right|_{\sigma_{ATM}} \\
\left.\frac{\partial^2 V^{BS}}{\partial \sigma^2}\right|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \left.\frac{\partial^2 U_k^{BS}}{\partial \sigma^2}\right|_{\sigma_{ATM}} \\
\left.\frac{\partial^2 V^{BS}}{\partial \sigma \partial S}\right|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \left.\frac{\partial^2 U_k^{BS}}{\partial \sigma \partial S}\right|_{\sigma_{ATM}}
\end{aligned} \tag{5.2.15}$$

This method was invented and adopted by FX traders based on empirical understanding, and is almost exclusively limited to the FX world. An obvious reason that this has largely been an FX tool may be because only in FX do we see so few strikes available; a

consistent pricing method will need to agree with the market values on a limited set of vol pillars. The price correction $\sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS} \right)$ is often called the "overhedge" [104] in banks.

Assume that $k = 1$ denotes the 25Δ put strike, $k = 2$ denotes the ATM strike, and $k = 3$ denotes the 25Δ call strike. In FX trading, the flat vol used for V^{BS} and U_k^{BS} is frequently updated to the current ATM market implied vol, hence $U_2^{mkt} = U_2^{BS}$. In this case, we have

$$V^{VV} = V^{BS} + \omega_1 \left(U_1^{mkt} - U_1^{BS}\big|_{\sigma_{ATM}} \right) + \omega_3 \left(U_3^{mkt} - U_3^{BS}\big|_{\sigma_{ATM}} \right) \tag{5.2.16}$$

If V is the price of a vanilla option with the price $U(K)$, Castagna and Mercurio [16] gives the unique solution to (5.2.15),

$$\omega_1 = \frac{\nu(K) \ln \frac{K_2}{K} \ln \frac{K_3}{K}}{\nu(K_1) \ln \frac{K_2}{K_1} \ln \frac{K_3}{K_1}},\ \omega_2 = \frac{\nu(K) \ln \frac{K}{K_1} \ln \frac{K_3}{K}}{\nu(K_2) \ln \frac{K_2}{K_1} \ln \frac{K_3}{K_2}},\ \omega_3 = \frac{\nu(K) \ln \frac{K}{K_1} \ln \frac{K}{K_2}}{\nu(K_3) \ln \frac{K_3}{K_1} \ln \frac{K_3}{K_2}} \tag{5.2.17}$$

where

$$\nu(K) = \frac{\partial U^{BS}}{\partial \sigma}(K) = S_0 e^{-r_f T} \sqrt{T} \varphi(d_1(K)) \tag{5.2.18}$$

and

$$d_1(K) = \frac{\ln \frac{S_0}{K} + \left(r_d - r_f + \frac{1}{2}\sigma^2 \right) T}{\sigma \sqrt{T}} \tag{5.2.19}$$

In the above formula, φ is the normal density function, r_d the domestic interest rate, and r_f the foreign interest rate.

Due to the analysis given by Castagna and Mercurio to this formula of the VV method (5.2.14), we name it the Castagna and Mercurio representation so as to differentiate this representation from other representations which will be discussed later.

The minimum consistency test is to see whether this pricing method will agree on the market prices of three plain vanillas. This agreement can be easily verified by the following VV prices on the three liquid strikes:

$$\begin{aligned}
U^{VV}(K_1) &= U^{BS}(K_1)\big|_{\sigma_{ATM}} + 1 \times \left(U^{mkt}(K_1) - U^{BS}(K_1)\big|_{\sigma_{ATM}} \right) = U^{mkt}(K_1) \\
U^{VV}(K_2) &= U^{BS}(K_2)\big|_{\sigma_{ATM}} + 1 \times \left(U^{mkt}(K_2) - U^{BS}(K_2)\big|_{\sigma_{ATM}} \right) = U^{mkt}(K_3) \\
U^{VV}(K_3) &= U^{BS}(K_3)\big|_{\sigma_{ATM}} + 1 \times \left(U^{mkt}(K_3) - U^{BS}(K_3)\big|_{\sigma_{ATM}} \right) = U^{mkt}(K_3)
\end{aligned} \tag{5.2.20}$$

where $U^{VV}(K)$ means the VV price at K. The above equalities hold because we can tell from (5.2.17) that $\omega_i(K_j)$ is 1 when $i = j$, zero when $i \neq j$.

Although the VV method is empirically found to be useful, it has not been adequately studied. Castagna and Mercurio [15, 16] showed that the VV method produces results rather close to those of the SABR [40] model for the smile between the 10Δ put and 10Δ call. Castagna and Mercurio also attempted to provide an explanation for the VV method. Castagna and Mercurio constructed a hedged portfolio consisting of a target vanilla instrument (so $V = U(K)$ in (5.2.14)), ΔS amount of currency, and a portfolio of three liquid vanilla options which is expressed by $\sum_{k=1}^{3} \omega_k U_k$. In trading, ΔS is used to hedge the Delta risk and the portfolio of vanillas at the vol pillars are used to hedge the Vega risk. Suppose at a particular time all of the option prices are valued by the Black-Scholes formula. However, the implied volatility is stochastic so the Black-Scholes prices are constantly revalued to the new volatility as the time t progresses. When there are changes in the option price $U(K)$, exchange rate S, and vanilla prices U_k, the hedged portfolio will change in value. The change in value of the hedged portfolio is expressed by Itô's Lemma:

$$
\begin{aligned}
& dV^{BS} - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{BS} \\
&= dU^{BS}(K) - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{BS} \\
&= \left[\frac{\partial U^{BS}(K)}{\partial t} - \sum_{k=1}^{3} \omega_k \frac{\partial U_k^{BS}}{\partial t} \right] dt \\
&+ \left[\frac{\partial U^{BS}(K)}{\partial S} - \Delta - \sum_{k=1}^{3} \omega_k \frac{\partial U_k^{BS}}{\partial S} \right] dS \\
&+ \left[\frac{\partial U^{BS}(K)}{\partial \sigma} - \sum_{k=1}^{3} \omega_k \frac{\partial U_k^{BS}}{\partial \sigma} \right] d\sigma \\
&+ \frac{1}{2} \left[\frac{\partial^2 U^{BS}(K)}{\partial S^2} - \sum_{k=1}^{3} \omega_k \frac{\partial^2 U_k^{BS}}{\partial S^2} \right] (dS)^2 \\
&+ \frac{1}{2} \left[\frac{\partial^2 U^{BS}(K)}{\partial \sigma^2} - \sum_{k=1}^{3} \omega_k \frac{\partial^2 U_k^{BS}}{\partial \sigma^2} \right] (d\sigma)^2 \\
&+ \left[\frac{\partial^2 U^{BS}(K)}{\partial S \partial \sigma} - \sum_{k=1}^{3} \omega_k \frac{\partial^2 U_k^{BS}}{\partial S \partial \sigma} \right] dS d\sigma
\end{aligned}
\tag{5.2.21}
$$

where σ is the ATM Black-Scholes implied vol σ_{ATM} and the Black-Scholes prices and Greeks are computed at σ_{ATM}. By doing this expansion, the authors actually made an implicit assumption about σ_{ATM} being a function of variables following diffusion processes, which we will discuss later.

Itô's Lemma is used instead of Taylor's expansion because here dt, dS, and $d\sigma$ conceptually represent a step of stochastic movement from the present time. When S is assumed to follow a diffusion process and the Black-Scholes implied volatility σ is assumed to be a function of diffusion processes, the higher-than-second-order terms will be zero in the statistical sense; this will be further discussed a bit later. The purpose of (5.2.21) is to pick the weights ω_k and Δ so that the expectation of the future growth of $dV^{BS} - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{BS}$ is zero. In reality, however, continuous hedging in (5.2.21) is not possible, and discrete hedging takes place instead. The higher-than-second-order terms may not be zero in the discrete hedging case; but as these terms are in higher order, they are usually much smaller than the lower order terms. We can also say that (5.2.21) is a truncated Taylor's expansion, which appears to be in the same form as the Itô's expansion. If (5.2.21) was regarded a truncated Taylor's expansion, almost all of our late analysis would be the same. In the analysis of the VV method in this chapter, however, we agree with Castagna and Mercurio in using the Itô's Lemma because we want to emphasise that the VV price is conceptually based on choosing the hedging strategies so as to achieve the zero expectation of hedging error in the Black-Scholes world looking into the future period dt immediately after the present time t. In a later chapter on the Karasinski method, we will use the Taylor's expansion in the analysis because the Karasinski method will be assumed to have no model error and have only parameter value estimation errors at the present time t. By contrast, the VV method has no parameter value estimation errors because S and σ_{ATM} are observable in the market; but the VV method has a model error (the flat-vol Black-Scholes cannot be right for a market with a smile), hence a hedging argument applies to the VV method. This subtle difference can later be seen in the barrier option example where a P factor is needed in the VV method but not in the Karasinski method.

The authors suggest that this expansion about σ_{ATM} may be thought of carrying contradiction because the Black-Scholes implied vol is usually viewed as a constant. But this can be justified because FX traders habitually run their books by hedging with the flat vol Black-Scholes model, and continuously updating the at-the-money volatility to the market level. Therefore, $U^{BS}(K)$ and U_k^{BS} can be simply regarded as functions of the stochastic ATM volatility, and the values of these functions happen to equal the stochastic

price movements of these instruments. Therefore, this analysis captures the essence of traders' hedging and there is no contradiction.

Castagna and Mercurio argue that after the stochastic components are balanced out by choosing the VV weights ω_k (5.2.15) and Δ for (5.2.21), $dU^{BS}(K) - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{BS}$ will have to grow at the risk free rate:

$$\begin{aligned} & dU^{BS}(K) - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{BS} \\ & = r_d \left[U^{BS}(K) - \Delta S - \sum_{k=1}^{3} \omega_k U_k^{BS} \right] dt \end{aligned} \tag{5.2.22}$$

Note that the above equality only makes sense in the Black-Scholes model values and it does not mean that the market value of the portfolio $dU^{mkt}(K) - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{mkt}$ will grow at the risk free rate.

Castagna and Mercurio made an example to illustrate how the hedging will perform in market reality. We will make it a bit more general. (5.2.22) can be re-arranged into

$$\begin{aligned} & d\left[U^{BS}(K) + \sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS} \right) \right] - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{mkt} \\ & = r_d \left[U^{BS}(K) + \sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS} \right) - \Delta S - \sum_{k=1}^{3} \omega_k U_k^{mkt} \right] dt \end{aligned}$$

hence

$$\begin{aligned} & dU^{VV} - \Delta dS - \sum_{k=1}^{3} \omega_k dU_k^{mkt} \\ & = r_d \left[U^{VV} - \Delta S - \sum_{k=1}^{3} \omega_k U_k^{mkt} \right] dt \end{aligned}$$

If the VV mark-to-market price U^{VV} is indeed the market price, ω_k will make a perfect hedge. In reality, it is impossible to have a perfect hedge because the VV price of an option V^{VV} cannot be a perfect duplicate of the unobservable market price V^{mkt}.

5.2.2.1. *Model Consistency of the VV Method with the Liquid Prices U_k^{mkt}.* A fundamental requirement of a model is the ability of fitting the liquid market vanilla prices U_k^{mkt}. Castagna and Mercurio [16] gave the consistency proof for the VV method based on the analytical solutions of ω_k using the Black-Scholes formula and the specific combination of sensitivities Vega, Volga and Vanna. We will show the VV method keeps the required price consistency on the liquid strikes regardless of the choices of sensitivities.

Assume that ω_k are computed with three sensitivities, whose operators are defined by D_1, D_2 and D_3. Each D is not limited to Vega, Volga or Vanna; instead it can be any sensitivity allowed by the Black-Scholes model.

Without loss of generality, assume that we try to price the vanilla U_1 with the VV method. Hence (5.2.15) becomes

$$\begin{aligned} D_1U_1^{BS}\big|_{\sigma_{ATM}} &= \omega_1\, D_1U_1^{BS}\big|_{\sigma_{ATM}} + \omega_2\, D_1U_2^{BS}\big|_{\sigma_{ATM}} + \omega_3\, D_1U_3^{BS}\big|_{\sigma_{ATM}} \\ D_2U_1^{BS}\big|_{\sigma_{ATM}} &= \omega_1\, D_2U_1^{BS}\big|_{\sigma_{ATM}} + \omega_2\, D_2U_2^{BS}\big|_{\sigma_{ATM}} + \omega_3\, D_2U_3^{BS}\big|_{\sigma_{ATM}} \\ D_3U_1^{BS}\big|_{\sigma_{ATM}} &= \omega_1\, D_3U_1^{BS}\big|_{\sigma_{ATM}} + \omega_2\, D_3U_2^{BS}\big|_{\sigma_{ATM}} + \omega_3\, D_3U_3^{BS}\big|_{\sigma_{ATM}} \end{aligned} \tag{5.2.23}$$

to which $\omega_1 = 1$, $\omega_2 = 0$, $\omega_3 = 0$ is obviously a solution.

Assume the matrix

$$A = \begin{pmatrix} D_1U_1^{BS}\big|_{\sigma_{ATM}} & D_1U_2^{BS}\big|_{\sigma_{ATM}} & D_1U_3^{BS}\big|_{\sigma_{ATM}} \\ D_2U_1^{BS}\big|_{\sigma_{ATM}} & D_2U_2^{BS}\big|_{\sigma_{ATM}} & D_2U_3^{BS}\big|_{\sigma_{ATM}} \\ D_3U_1^{BS}\big|_{\sigma_{ATM}} & D_3U_2^{BS}\big|_{\sigma_{ATM}} & D_3U_3^{BS}\big|_{\sigma_{ATM}} \end{pmatrix} \tag{5.2.24}$$

is invertible; under this assumption, the weights ω_k admit a unique solution. Under this assumption, $\omega_1 = 1$, $\omega_2 = 0$, $\omega_3 = 0$ is the unique solution. Therefore, the VV price for U_1 is $U_1^{model} + 1\left(U_1^{mkt} - U_1^{model}\right) = U_1^{mkt}$, a perfect fit of the market price of the liquid vanilla. For the other vanillas U_2 and U_3, the same proof applies.

From the above analysis, we can see what types of sensitivities (Vega, Volga, or Vanna) do not matter much in fitting the market vanilla prices on the liquid strikes with the VV formula. What is important is that the number of liquid vanillas is equal to the number of sensitivities so the solution $\omega_{k'} = 1$, $\omega_{k\neq k'} = 0$ can be obtained for a liquid vanilla $U_{k'}$.

Although the error of the VV formula on the liquid vanillas is zero, the combination of sensitivities used in computation of ω_k is not arbitrary; it has an effect on the error on the prices of options on strikes that are not liquid. We will analyse the error between the VV and market prices under different market assumptions later. We will use a test to show how the choice of sensitivities affects the error.

5.2.3. Implicit Assumptions in Castagna and Mercurio's Hedging Analysis.

5.2.3.1. *Option Prices Depending Only on Current Spot.* An obvious assumption for the Castagna and Mercurio's hedging analysis (5.2.21) is that the option prices only depend on current spot as far as the spot is concerned. If the target option V is a path dependent

option whose price is based on a history of spot prices, there will be extra terms in (5.2.21). In the similar analysis in the following text of this thesis, we will make the same assumption unless specified otherwise.

5.2.3.2. *σ_{ATM} being a function of variables following diffusion processes.* By taking the Itô's expansion in (5.2.21), Castagna and Mercurio actually made an implicit assumption about the process of the ATM Black-Scholes implied volatility σ_{ATM} being a function of stochastic variables following diffusion processes. For example, if the market follows the CEV model the Black-Scholes implied vol depends only on S; if the market follows the Heston model, the Black-Scholes implied vol depends on S and ν; if the market follows the double-Heston model, the Black-Scholes implied vol depends on S, ν_1 and ν_2. So σ_{ATM} is assumed to follow

$$\sigma_{ATM} = \sigma_{ATM}\left(S,\, X_1,\, X_2, ...,\, X_n, ...X_N\right) \tag{5.2.25}$$

where each of the stochastic variables $X_n,\ 1 \leqslant n \leqslant N$, follows a diffusion process. So the Itô's expansion of σ_{ATM} will take the form of

$$d\sigma_{ATM} = (...)\, dt + \sum_{n=1}^{N} (...)\, dW_n \tag{5.2.26}$$

where dW_n, $1 \leqslant n \leqslant N$, are the Wiener processes.

5.2.3.3. *Hedging with Three Vanillas or One?* In the hedging of the portfolio in (5.2.21), the aim is to remove the stochasticity of the portfolio, It is easy to see that the coefficients of $d\sigma$, $(d\sigma)^2$, and $dSd\sigma$ will be zero when (5.2.15) holds. Hence making the coefficients of $d\sigma$, $(d\sigma)^2$, and $dSd\sigma$ in (5.2.21) all zero would require three vanillas.

However, because σ is assumed to follow the process (5.2.26), $(d\sigma)^2$ and $dSd\sigma$ can be rearranged to become part of the dt term. It appears that holding ω units of just one hedging vanilla would be needed to cancel out the coefficient of the $d\sigma$ term. The Itô's

expansion of the portfolio of an option V , some currency and one hedging vanilla U gives

$$\begin{aligned}
&dV^{BS} - \Delta dS - \omega dU^{BS} \\
&= \left[\frac{\partial V^{BS}}{\partial t} - \omega\frac{\partial U^{BS}}{\partial t}\right] dt \\
&+ \left[\frac{\partial V^{BS}}{\partial S} - \Delta - \omega\frac{\partial U^{BS}}{\partial S}\right] dS \\
&+ \left[\frac{\partial V^{BS}}{\partial \sigma} - \omega\frac{\partial U^{BS}}{\partial \sigma}\right] d\sigma \\
&+ \frac{1}{2}\left[\frac{\partial^2 V^{BS}}{\partial S^2} - \omega\frac{\partial^2 U^{BS}}{\partial S^2}\right] (...)\, dt \\
&+ \frac{1}{2}\left[\frac{\partial^2 V^{BS}}{\partial \sigma^2} - \omega\frac{\partial^2 U^{BS}}{\partial \sigma^2}\right] (...)\, dt \\
&+ \left[\frac{\partial^2 V^{BS}}{\partial S \partial \sigma} - \omega\frac{\partial^2 U^{BS}}{\partial S \partial \sigma}\right] (...)\, dt
\end{aligned} \tag{5.2.27}$$

where σ is the ATM Black-Scholes implied vol σ_{ATM} and V^{BS}, U^{BS} and their Greeks are evaluated at σ_{ATM}.

If we have the perfect knowledge of (5.2.25), using one vanilla to hedge is quite enough, by making $\dfrac{\partial V^{BS}}{\partial \sigma} = \omega\dfrac{\partial U^{BS}}{\partial \sigma}$. There are three practical problems with this approach. First, the relationship (5.2.25) is not directly market observable. One may calibrate (5.2.25) based on one's view on the market. (5.2.25) depends on the model of choice. But there is no market consensus on what specific model the market is exactly following and what are the parameters for this model. So it is likely there is always imperfection in the understanding of the function and parameters of (5.2.25). Second, even if our choice of model is perfectly following the market, it is not realistic to expect continuous calibration to make the model parameters perfect in tune with a fast changing market. Finally, as we have discussed earlier, the price correction formula based on only one vanilla instrument will not be able to fit all the market prices on three liquid strikes.

In light of the practical issues, a simple way of removing the concern over σ_{ATM} movement is through making the coefficients of not only $d\sigma$ but also $(d\sigma)^2$ and $dSd\sigma$ zero. This approach does not require the knowledge of the specific model and parameters of the σ process at all. But with only one hedging vanilla, it is impossible because just one ω cannot possibly satisfy three equalities

$$\frac{\partial V^{BS}}{\partial \sigma} = \omega\frac{\partial U^{BS}}{\partial \sigma},\ \frac{\partial^2 V^{BS}}{\partial \sigma^2} = \omega\frac{\partial^2 U^{BS}}{\partial \sigma^2},\ \frac{\partial^2 V^{BS}}{\partial S \partial \sigma} = \omega\frac{\partial^2 U^{BS}}{\partial S \partial \sigma}$$

where the above Greeks of V^{BS} are evaluated at σ_{ATM}.

That is why the VV method demands three vanilla instruments. The FX market seems to be lucky enough to happen to have three liquid vanillas. This perhaps explains why the VV method was invented in the FX trading community. With three liquid vanilla options, the VV weights computed by (5.2.15) cause the uncertainty in (5.2.27) to go away.

So in the hedging analysis in this thesis later on, we will always choose to cancel out $d\sigma$, $(d\sigma)^2$ and $dSd\sigma$ terms unless otherwise specified.

Note in Chapter 7, we will introduce another method called the Karasinski method, which is based on using the right market model with some slightly wrong parameters. In the discussion of the Karasinski method, which has no model error but has a parameter uncertainty error, the focus is to remove the parameter uncertainty. However, there is hardly any parameter uncertainty in the VV method, for which the model parameter σ_{ATM} is market observable. With a model error but no parameter uncertainty error, the VV method is based on a hedging argument whereas the Karasinski method will be shown more appropriate to follow a parameter error argument. We will further compare these two methods later in the thesis.

5.2.4. Alternative Castagna and Mercurio Form. It was pointed out by a referee of the Castagna and Mercurio [16] paper that the VV price can alternatively be expressed by the sum of the Black-Scholes price, and a linear combination of Vega, Volga, and Vanna of the Black-Scholes price.

Define

$$C = \begin{pmatrix} U_1^{mkt} - U_1^{BS} \\ U_2^{mkt} - U_2^{BS} \\ U_3^{mkt} - U_3^{BS} \end{pmatrix} \qquad B = \begin{pmatrix} V_\sigma^{BS} \\ V_{\sigma\sigma}^{BS} \\ V_{\sigma S}^{BS} \end{pmatrix}\Bigg|_{\sigma_{ATM}}$$
$$A = \begin{pmatrix} \left(U_1^{BS}\right)_\sigma & \left(U_2^{BS}\right)_\sigma & \left(U_3^{BS}\right)_\sigma \\ \left(U_1^{BS}\right)_{\sigma\sigma} & \left(U_2^{BS}\right)_{\sigma\sigma} & \left(U_3^{BS}\right)_{\sigma\sigma} \\ \left(U_1^{BS}\right)_{\sigma S} & \left(U_2^{BS}\right)_{\sigma S} & \left(U_3^{BS}\right)_{\sigma S} \end{pmatrix}\Bigg|_{\sigma_{ATM}} \qquad \omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{5.2.28}$$

The VV price gives

$$\begin{aligned} V^{VV} &= V^{BS}\big|_{\sigma_{ATM}} + C' \cdot \omega \\ B &= A \cdot \omega \end{aligned} \tag{5.2.29}$$

hence

$$V^{VV} = V^{BS}\big|_{\sigma_{ATM}} + C' \cdot A^{-1} \cdot B \tag{5.2.30}$$

Define $Y = C^{'} \cdot A^{-1} = (y_1,\ y_2\ , y_3)$. (5.2.30) can be written as

$$\begin{aligned} V^{VV} &= V^{BS}\big|_{\sigma_{ATM}} + Y \cdot B \\ &= V^{BS}\big|_{\sigma_{ATM}} + y_1\ V^{BS}_{\sigma}\big|_{\sigma_{ATM}} + y_2\ V^{BS}_{\sigma\sigma}\big|_{\sigma_{ATM}} + y_3\ V^{BS}_{\sigma S}\big|_{\sigma_{ATM}} \end{aligned} . \tag{5.2.31}$$

The above Y is independent of $V^{BS}\big|_{\sigma_{ATM}}$. Therefore, it can be computed just once, which is a major benefit of using this representation. This expression of V^{VV} may be viewed as a combination of the ATM vol price, hedging costs of Vega, Vanna, and Volga at the ATM vol. The process of computing Y can equally be viewed as a simple calibration based on three market vanillas.

5.2.5. Conceptual Difficulties with the VV Method as a Hedging Tool. Assume a portfolio consisting of an option whose unobservable market price is denoted by V^{mkt}, some currency ΔS, and a portfolio of three liquid vanillas with the sum of market prices $\sum\limits_k \omega_k U_k^{mkt}$. A change of market portfolio value driven by the changes in the underlying price and volatility follows

$$\begin{aligned} & dV^{mkt} - \Delta dS - \sum_k \omega_k dU_k^{mkt} \\ &= dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS} + d\left(V^{mkt} - V^{VV}\right) \end{aligned} \tag{5.2.32}$$

where V^{BS} and U_k^{BS} denote the Black-Scholes prices valued at σ_{ATM}, and V^{VV} denotes the VV mark-to-market price.

The above equality holds when the mark-to-market price V^{VV} is chosen to be the VV price described in (5.2.14), which can also be re-arranged into

$$\begin{aligned} & dV^{VV} - \Delta dS - \sum_k \omega_k dU_k^{mkt} \\ &= dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS} \end{aligned} \tag{5.2.33}$$

Expanding $dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS}$ by Itô's Lemma gives

$$\begin{aligned}
&dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS} \\
&= \left[\frac{\partial V^{BS}}{\partial t} - \sum_k \omega_k \frac{\partial U_k^{BS}}{\partial t}\right] dt + \left[\frac{\partial V^{BS}}{\partial S} - \Delta - \sum_k \omega_k \frac{\partial U_k^{BS}}{\partial S}\right] dS \\
&+ \frac{1}{2}\left[\frac{\partial^2 V^{BS}}{\partial S^2} - \sum_k \omega_k \frac{\partial^2 U_k^{BS}}{\partial S^2}\right] S^2\sigma^2 dt \\
&= \left[\frac{\partial V^{BS}}{\partial t} + \frac{1}{2}S^2\sigma^2\frac{\partial^2 V^{BS}}{\partial S^2} - \sum_k \omega_k \left(\frac{\partial U_k^{BS}}{\partial t} + \frac{1}{2}S^2\sigma^2\frac{\partial^2 U_k^{BS}}{\partial S^2}\right)\right] dt \\
&+ \left[\frac{\partial V^{BS}}{\partial S} - \Delta - \sum_k \omega_k \frac{\partial U_k^{BS}}{\partial S}\right] dS
\end{aligned} \tag{5.2.34}$$

where the $d\sigma$, $(d\sigma)^2$, $dSd\sigma$ terms have been cancelled out with the choices of the VV weights ω_k computed through the Vega, Volga, and Vanna neutrality in (5.2.15), and the Black-Scholes prices and Greeks are computed with σ_{ATM}.

Choosing the Delta hedge

$$\Delta = \frac{\partial V^{BS}}{\partial S} - \sum_k \omega_k \frac{\partial U_k^{BS}}{\partial S}, \tag{5.2.35}$$

we will make the coefficient of dS_t zero. From the fact that V^{BS} and U_k^{BS} satisfy the Black-Scholes PDE

$$\frac{\partial u}{\partial t} + \frac{1}{2}S^2\sigma^2\frac{\partial^2 u}{\partial S^2} + (r_d - r_f)\, S\frac{\partial u}{\partial S} - r_d u = 0, \tag{5.2.36}$$

we obtain

$$\begin{aligned}
&\frac{\partial V^{BS}}{\partial t}+\frac{1}{2}S^2\sigma^2\frac{\partial^2 V^{BS}}{\partial S^2}\\
&-\sum_k \omega_k\left(\frac{\partial U_k^{BS}}{\partial t}+\frac{1}{2}S^2\sigma^2\frac{\partial^2 U_k^{BS}}{\partial S^2}\right)\\
&=r_d V^{BS}-\left(r_d-r_f\right)S\frac{\partial V^{BS}}{\partial S}\\
&-\sum_{k=1}^{3}\omega_k\left(r_d U_k^{BS}-\left(r_d-r_f\right)S\frac{\partial U_k^{BS}}{\partial S}\right)\\
&=r_d\left[V^{BS}-\sum_k \omega_k U_k^{BS}\right]\\
&-\left(r_d-r_f\right)S\left[\frac{\partial V^{BS}}{\partial S}-\sum_k \omega_k\frac{\partial U_k^{BS}}{\partial S}\right]\\
&=r_d\left[V^{BS}-\sum_k \omega_k U_k^{BS}\right]-\left(r_d-r_f\right)S\Delta\\
&=r_d\left[V^{BS}-S\Delta-\sum_k \omega_k U_k^{BS}\right]+r_f S\Delta
\end{aligned}\tag{5.2.37}$$

Combining (5.2.37), (5.2.34) and (5.2.33), we obtain

$$\begin{aligned}
&dV^{BS}-\Delta dS-\sum_k \omega_k dU_k^{BS}\\
&=r_d\left[V^{BS}-S\Delta-\sum_k \omega_k U_k^{BS}\right]dt+r_f S\Delta dt
\end{aligned}\tag{5.2.38}$$

With the VV formula (5.2.14), (5.2.38) can be changed to

$$\begin{aligned}
&dV^{VV}-\Delta dS-\sum_k \omega_k dU_k^{mkt}\\
&=r_d\left[V^{VV}-S\Delta-\sum_k \omega_k U_k^{mkt}\right]dt+r_f S\Delta dt
\end{aligned}\tag{5.2.39}$$

and

$$\begin{aligned}
&dV^{mkt}-\Delta dS-\sum_k \omega_k dU_k^{mkt}\\
&=r_d\left[V^{mkt}-S\Delta-\sum_k \omega_k U_k^{mkt}\right]dt+r_f S\Delta dt\\
&\quad+d\left(V^{mkt}-V^{VV}\right)-r_d\left(V^{mkt}-V^{VV}\right)dt
\end{aligned}\tag{5.2.40}$$

Here we see the hedged portfolio locally grows at the rate of r_d, the domestic risk free rate, plus $d\left(V^{mkt}-V^{VV}\right)-r_d\left(V^{mkt}-V^{VV}\right)dt$. There is also an $r_f S\Delta dt$ term. To explain this term, we notice the similarity between (5.2.36) and the standard Black-Scholes equation with dividend payments. The foreign rate r_f corresponds to the dividend rate. $r_f S\Delta dt$ means the initial cash $S\Delta$ grows at the rate of r_f, which can be viewed

as the dividend delivered by the foreign interest. The component $d\left(V^{mkt}-V^{VV}\right)$ is still stochastic; therefore the risk has not been fully hedged away.

If the VV price V^{VV} follows V^{mkt} exactly and $V^{mkt}-V^{VV}$ grows at the rate r_d, the market portfolio in (5.2.40) will always grow at the rate r_d. However, the mark-to-market price V cannot possibly be a perfect copy of V^{mkt}, so a perfect hedge is impossible, making the VV method not a perfect arbitrage-free hedging tool. However, as the VV price is considered a very good mark-to-market price, in particularly for vanillas, the difference $d\left(V^{mkt}-V^{VV}\right)$ is usually not large. We will demonstrate this using numerical analysis on the performance of the VV price later.

Therefore, the above analysis shows that the VV method is not a perfect hedging tool theoretically; it is not possible to ascertain that the VV method is free of arbitrage. However, we will show in a hedging test with the VV strategies later in this chapter that the VV method can be an excellent hedging tool with very low hedging error and very small standard deviation of hedging error in a stochastic simulated market.

5.2.6. The VV Method Expressed by Risk Reversal and Butterfly. If we use the risk reversal $\widetilde{RR}=U_3-U_1$ and the butterfly $\widetilde{BF}=\left(U_3-2U_2+U_1\right)/2$ (both are defined in (5.1.9)) to replace the vanillas U_1 and U_3, respectively, the VV price (denoted by $V^{RRBF-VV}$ in this representation) will take the form of

$$\begin{aligned} V^{RRBF-VV} = V^{BS}\big|_{\sigma_{ATM}} + \omega_{\widetilde{RR}}\left(\widetilde{RR}^{mkt} - \widetilde{RR}^{BS}\Big|_{\sigma_{ATM}}\right) + \\ \omega_{ATM}\left(U_{ATM}^{mkt} - U_{ATM}^{BS}\big|_{\sigma_{ATM}}\right) + \omega_{\widetilde{BF}}\left(\widetilde{BF}^{mkt} - \widetilde{BF}^{BS}\Big|_{\sigma_{ATM}}\right) \end{aligned} \tag{5.2.41}$$

where

$$\left.\begin{pmatrix} \dfrac{\partial V^{BS}}{\partial\sigma} \\ \dfrac{\partial^2 V^{BS}}{\partial\sigma^2} \\ \dfrac{\partial^2 V^{BS}}{\partial S\partial\sigma} \end{pmatrix}\right|_{\sigma_{ATM}} = \left.\begin{pmatrix} \dfrac{\partial\widetilde{RR}^{BS}}{\partial\sigma} & \dfrac{\partial U_{ATM}^{BS}}{\partial\sigma} & \dfrac{\partial\widetilde{BF}^{BS}}{\partial\sigma} \\ \dfrac{\partial^2\widetilde{RR}^{BS}}{\partial\sigma^2} & \dfrac{\partial^2 U_{ATM}^{BS}}{\partial\sigma^2} & \dfrac{\partial^2\widetilde{BF}^{BS}}{\partial\sigma^2} \\ \dfrac{\partial^2\widetilde{RR}^{BS}}{\partial S\partial\sigma} & \dfrac{\partial^2 U_{ATM}^{BS}}{\partial S\partial\sigma} & \dfrac{\partial^2\widetilde{BF}^{BS}}{\partial S\partial\sigma} \end{pmatrix}\right|_{\sigma_{ATM}} \begin{pmatrix} \omega_{\widetilde{RR}} \\ \omega_{ATM} \\ \omega_{\widetilde{BF}} \end{pmatrix} \tag{5.2.42}$$

Remember that U_2 is U_{ATM}. $\widetilde{RR}$ and $\widetilde{BF}$ are composed of U_1, U_2 and U_3. They can be viewed as liquid prices. According the analysis for the VV method, $V^{RRBF-VV}$ defined by (5.2.41) will produce the market prices U_{ATM}^{mkt}, $\widetilde{RR}^{mkt}$ and $\widetilde{BF}^{mkt}$; therefore, the market prices U_1^{mkt}, U_2^{mkt} and U_3^{mkt} can be produced by this VV formula.

We will now prove that the $V^{RRBF-VV}$ price is just a different form of the V^{VV} price. Both prices are actually equal to each other. By the definitions of $\widetilde{RR}$ and $\widetilde{BF}$, (5.2.41) can be rearranged to

$$V^{RRBF-VV} = V^{BS}\big|_{\sigma_{ATM}} + \sum_{k=1}^{3} \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \tag{5.2.43}$$

where

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} -1 & 0 & \frac{1}{2} \\ 0 & 1 & -1 \\ 1 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \omega_{\widetilde{RR}} \\ \omega_{ATM} \\ \omega_{\widetilde{BF}} \end{pmatrix} \tag{5.2.44}$$

Equations (5.2.44) give

$$\begin{pmatrix} \omega_{\widetilde{RR}} \\ \omega_{ATM} \\ \omega_{\widetilde{BF}} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{5.2.45}$$

Combining (5.2.42) and (5.2.45) gives

$$\begin{aligned} \begin{pmatrix} \frac{\partial V^{BS}}{\partial \sigma} \\ \frac{\partial^2 V^{BS}}{\partial \sigma^2} \\ \frac{\partial^2 V^{BS}}{\partial S \partial \sigma} \end{pmatrix}\Bigg|_{\sigma_{ATM}} &= \begin{pmatrix} \frac{\partial \widetilde{RR}^{BS}}{\partial \sigma} & \frac{\partial U_{ATM}^{BS}}{\partial \sigma} & \frac{\partial \widetilde{BF}^{BS}}{\partial \sigma} \\ \frac{\partial^2 \widetilde{RR}^{BS}}{\partial \sigma^2} & \frac{\partial^2 U_{ATM}^{BS}}{\partial \sigma^2} & \frac{\partial^2 \widetilde{BF}^{BS}}{\partial \sigma^2} \\ \frac{\partial^2 \widetilde{RR}^{BS}}{\partial S \partial \sigma} & \frac{\partial^2 U_{ATM}^{BS}}{\partial S \partial \sigma} & \frac{\partial^2 \widetilde{BF}^{BS}}{\partial S \partial \sigma} \end{pmatrix}\Bigg|_{\sigma_{ATM}} \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial U_1^{BS}}{\partial \sigma} & \frac{\partial U_2^{BS}}{\partial \sigma} & \frac{\partial U_3^{BS}}{\partial \sigma} \\ \frac{\partial^2 U_1^{BS}}{\partial \sigma^2} & \frac{\partial^2 U_2^{BS}}{\partial \sigma^2} & \frac{\partial^2 U_3^{BS}}{\partial \sigma^2} \\ \frac{\partial^2 U_1^{BS}}{\partial S \partial \sigma} & \frac{\partial^2 U_2^{BS}}{\partial S \partial \sigma} & \frac{\partial^2 U_3^{BS}}{\partial S \partial \sigma} \end{pmatrix}\Bigg|_{\sigma_{ATM}} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \end{aligned}$$

Therefore, $V^{RRBF-VV} = V^{VV}$ always holds. The formula of $V^{RRBF-VV}$ is just a different representation of the VV price V^{VV} defined in (5.2.14).

Because of $U_{ATM}^{mkt} = U_{ATM}^{BS}\big|_{\sigma_{ATM}}$, (5.2.41) can be written as

$$V^{RRBF-VV} = V^{BS}\big|_{\sigma_{ATM}} + \omega_{\widetilde{RR}} \left(\widetilde{RR}^{mkt} - \widetilde{RR}^{BS}\Big|_{\sigma_{ATM}} \right) + \omega_{\widetilde{BF}} \left(\widetilde{BF}^{mkt} - \widetilde{BF}^{BS}\Big|_{\sigma_{ATM}} \right) \tag{5.2.46}$$

where $\omega_{\widetilde{RR}}$ and $\omega_{\widetilde{BF}}$ satisfy (5.2.42).

For easy reference, we will name this VV representation the RRBF-VV method.

5.3. Traders' Rule Representation

Wystup [104] described a representation similar to the RRBF-VV method, which was also referred to as "the Traders' Rule of Thumb" (this representation will be called the "TR-VV" method in this thesis). This "Traders' Rule" representation of the TR-VV mark-to-market price (denoted by V^{TR-VV}) is expressed by

$$\begin{aligned} V^{TR-VV} &= V^{BS}\big|_{\sigma_{ATM}} + \omega_{\widetilde{RR}}\left(\widetilde{RR}^{mkt} - \widetilde{RR}^{BS}\Big|_{\sigma_{ATM}}\right) \\ &\quad + \omega_{\widetilde{BF}}\left(\widetilde{BF}^{mkt} - \widetilde{BF}^{BS}\Big|_{\sigma_{ATM}}\right) \end{aligned} \tag{5.3.1}$$

where

$$\begin{cases} \omega_{\widetilde{RR}} & \triangleq \dfrac{V^{mkt}_{\sigma\sigma}}{\widetilde{RR}^{mkt}_{\sigma S}},\ V^{mkt}_{\sigma\sigma}: \text{Black-Scholes Volga on Market Vol} \\ \omega_{\widetilde{BF}} & \triangleq \dfrac{V^{mkt}_{\sigma S}}{\widetilde{BF}^{mkt}_{\sigma\sigma}},\ V^{mkt}_{\sigma S}: \text{Black-Scholes Vanna on Market Vol} \\ \widetilde{RR}^{mkt}_{\sigma S} & \triangleq \left(U_3^{BS}\right)_{\sigma S}\big|_{\sigma_3} - \left(U_1^{BS}\right)_{\sigma S}\big|_{\sigma_1} \\ \widetilde{BF}^{mkt}_{\sigma\sigma} & \triangleq \dfrac{1}{2}\left(\left(U_1^{BS}\right)_{\sigma\sigma}\big|_{\sigma_1} + \left(U_3^{BS}\right)_{\sigma\sigma}\big|_{\sigma_3}\right) - \left(U_2^{BS}\right)_{\sigma\sigma}\big|_{\sigma_{ATM}} \end{cases} \tag{5.3.2}$$

where σ_1 is the market Black-Scholes implied vol for the strike of U_1, $\sigma_2 = \sigma_{ATM}$ is the market Black-Scholes implied vol for the strike of U_2, and σ_3 is the market Black-Scholes implied vol for the strike of U_3. $V^{mkt}_{\sigma\sigma}$ and $V^{mkt}_{\sigma S}$ are a bit tricky. They are priced on the Black-Scholes implied vol smile and not just on σ_{ATM}. Suppose we are trying to price $\widetilde{RR}$, hence $V^{mkt} = \widetilde{RR}^{mkt}$; according to Wystup [104], $\widetilde{RR}^{mkt}_{\sigma S}$ is defined above in (5.3.2) and $\widetilde{RR}^{mkt}_{\sigma\sigma} = \left(U_3^{BS}\right)_{\sigma\sigma}\big|_{\sigma_3} - \left(U_1^{BS}\right)_{\sigma\sigma}\big|_{\sigma_1}$. According to the above definitions, the estimates of $V^{mkt}_{\sigma\sigma}$ and $V^{mkt}_{\sigma S}$ for liquid strikes are straightforward because the market Black-Scholes implied vols are available on these strikes. But for other strikes or an exotic option with no strike parameter, it is unclear how $V^{mkt}_{\sigma\sigma}$ and $V^{mkt}_{\sigma S}$ can be exactly computed without accurate information on the market Black-Scholes implied vol for the instrument V; a simple approximation in practice is to just use the ATM vol to price them.

The TR-VV method also has a few assumptions [104], whose importance will be discussed later.

The TR-VV formula can be re-arranged into

$$\begin{aligned} V^{TR-VV} &\approx V^{BS}\big|_{\sigma_{ATM}} + \frac{\left(\widetilde{RR}^{mkt} - \widetilde{RR}^{BS}\big|_{\sigma_{ATM}}\right)}{\widetilde{RR}^{mkt}_{\sigma S}} V^{mkt}_{\sigma\sigma} + \frac{\left(\widetilde{BF}^{mkt} - \widetilde{BF}^{BS}\big|_{\sigma_{ATM}}\right)}{\widetilde{BF}^{mkt}_{\sigma\sigma}} V^{mkt}_{\sigma S} \\ &\triangleq V^{BS}\big|_{\sigma_{ATM}} + \text{Cost of Volga} + \text{Cost of Vanna} \end{aligned} \tag{5.3.3}$$

Comparing (5.3.3) to (5.2.31), (5.3.3) appears to lack an explicit "cost of Vega" term, which is the term containing $V^{BS}_{\sigma}\big|_{\sigma_{ATM}}$ in (5.2.31). However, it does not mean either representation is wrong. It is easy to verify that the cost-of-Volga term in (5.3.3) is not the same as the cost-of-Volga term in (5.2.31), and the cost-of-Vanna term in (5.3.3) not the same as the cost-of-Vanna term in (5.2.31). We will later see that this missing Vega is possible only when some assumptions are made.

Wystup [104] verified that the TR-VV method gives good fitting to the market prices of the liquid vanillas.

5.3.1. Assumptions of the Trader's Rule Representation. The TR-VV made the assumptions [104, 103] of zero Volga of the vanilla risk reversal, zero Vanna of vanilla butterfly, very small Volga and Vanna of ATM vanilla options:

$$\text{Assumptions}: \begin{cases} \dfrac{\partial^2 \widetilde{RR}^{mkt}}{\partial\sigma^2} \approx 0, \ \left.\dfrac{\partial^2 \widetilde{RR}^{BS}}{\partial\sigma^2}\right|_{\sigma_{ATM}} \approx 0 \\ \dfrac{\partial^2 \widetilde{BF}^{mkt}}{\partial S\partial\sigma} \approx 0, \ \left.\dfrac{\partial^2 \widetilde{BF}^{BS}}{\partial S\partial\sigma}\right|_{\sigma_{ATM}} \approx 0 \\ \dfrac{\partial^2 U^{mkt}_{ATM}}{\partial\sigma^2} \approx 0, \ \dfrac{\partial^2 U^{BS}_{ATM}}{\partial\sigma^2} \approx 0 \\ \dfrac{\partial^2 U^{mkt}_{ATM}}{\partial S\partial\sigma} \approx 0, \ \dfrac{\partial^2 U^{BS}_{ATM}}{\partial S\partial\sigma} \approx 0 \end{cases} \tag{5.3.4}$$

These assumptions hold in approximation in normal markets. But they are criticised because they are not exact [104, 103]. From a hedging error point of view, we will analyse the possible basis on which these assumptions are made.

A major difference between the TR-VV method and the RRBF-VV method is the way by which $\omega_{\widetilde{RR}}$ and $\omega_{\widetilde{BF}}$ are computed. The formula for $\omega_{\widetilde{RR}}$ and $\omega_{\widetilde{BF}}$ in the TR-VV method seems to be some approximate form of the formula in the RRBF-VV method. We will explore whether the assumptions (5.3.4) are linked to the specific formula to compute $\omega_{\widetilde{RR}}$ and $\omega_{\widetilde{BF}}$ in the TR-VV method.

Applying the assumptions (5.3.4) to the formula of the RRBF-VV formula (5.2.42) gives

$$\left.\begin{pmatrix} \dfrac{\partial V^{BS}}{\partial\sigma} \\ \dfrac{\partial^2 V^{BS}}{\partial\sigma^2} \\ \dfrac{\partial^2 V^{BS}}{\partial S\partial\sigma} \end{pmatrix}\right|_{\sigma_{ATM}} = \left.\begin{pmatrix} \dfrac{\partial \widetilde{RR}^{BS}}{\partial\sigma} & \dfrac{\partial U^{BS}_{ATM}}{\partial\sigma} & \dfrac{\partial \widetilde{BF}^{BS}}{\partial\sigma} \\ 0 & 0 & \dfrac{\partial^2 \widetilde{BF}^{BS}}{\partial\sigma^2} \\ \dfrac{\partial^2 \widetilde{RR}^{BS}}{\partial S\partial\sigma} & 0 & 0 \end{pmatrix}\right|_{\sigma_{ATM}} \begin{pmatrix} \omega_{\widetilde{RR}} \\ \omega_{ATM} \\ \omega_{\widetilde{BF}} \end{pmatrix}$$

hence

$$\begin{aligned} \omega_{\widetilde{RR}} &= \left.\frac{\partial^2 V^{BS}}{\partial S\partial\sigma}\right|_{\sigma_{ATM}} \Big/ \left.\frac{\partial^2 \widetilde{RR}^{BS}}{\partial S\partial\sigma}\right|_{\sigma_{ATM}} \\ \omega_{\widetilde{BF}} &= \left.\frac{\partial^2 V^{BS}}{\partial\sigma^2}\right|_{\sigma_{ATM}} \Big/ \left.\frac{\partial^2 \widetilde{BF}^{BS}}{\partial\sigma^2}\right|_{\sigma_{ATM}} \end{aligned}$$

Therefore, the mark-to-market price $V^{TRBS-VV}$ defined by

$$\begin{aligned} V^{TRBS-VV} \approx\ & V^{BS} + \left.\frac{V^{BS}_{S\sigma}}{\widetilde{RR}^{BS}_{S\sigma}}\right|_{\sigma_{ATM}} \left(\widetilde{RR}^{mkt} - \left.\widetilde{RR}^{BS}\right|_{\sigma_{ATM}}\right) \\ & + \left.\frac{V^{BS}_{\sigma\sigma}}{\widetilde{BF}^{BS}_{\sigma\sigma}}\right|_{\sigma_{ATM}} \left(\widetilde{BF}^{mkt} - \left.\widetilde{BF}^{BS}\right|_{\sigma_{ATM}}\right), \end{aligned} \tag{5.3.5}$$

is an approximation to the RRBF-VV price defined in (5.2.46). For easy reference, we name (5.3.5) the TRBS-VV method.

If we assume that U^{mkt}_{ATM}, $\widetilde{RR}^{mkt}$, and $\widetilde{BF}^{mkt}$ are twice differentiable, through the hedging analysis, we can similarly obtain the TR-VV price

$$V^{TR-VV} \approx V^{BS} + \frac{V^{mkt}_{S\sigma}}{\widetilde{RR}^{mkt}_{S\sigma}} \left(\widetilde{RR}^{mkt} - \left.\widetilde{RR}^{BS}\right|_{\sigma_{ATM}}\right) + \frac{V^{mkt}_{\sigma\sigma}}{\widetilde{BF}^{mkt}_{\sigma\sigma}} \left(\widetilde{BF}^{mkt} - \left.\widetilde{BF}^{BS}\right|_{\sigma_{ATM}}\right) \tag{5.3.6}$$

Our analysis shows that the TR-VV method and the TRBS-VV method are just approximated methods to the VV (or RRBF-VV) method, which requires no assumptions in (5.3.4). The computational costs are comparable between the VV method and the TR-VV (or TRBS-VV) method. Computing Greeks of the vanillas costs little time because of the existence of analytical solutions. Compared to the TR-VV method, the VV method will only need to compute an extra Vega of the target option V. Given current computing environment, the speed advantage gained in the TR-VV method does not seem to justify the need for an approximation to the VV price. As a result, in this thesis, we will focus on

the Castagna and Mercurio representation of the VV method defined in (5.2.14) and the corresponding Karasinski risk correction method.

5.4. Error Estimation of the Vanna-Volga Price

5.4.1. Reflection on the Hedging Arguments. In the above analysis about the VV price using the hedging arguments, assuming that the Black-Scholes implied volatility follows a diffusion process, we find the VV price V^{VV} in (5.2.14) and the VV ω_k in (5.2.15) will allow the portfolio value change $\delta V^{VV} - \Delta\delta S - \sum_k \omega_k \delta U_k^{mkt}$ to grow at risk-free rate. If the VV price V^{VV} is a perfect duplicate of the market price V^{mkt}, $V \equiv V^{mkt}$, $\delta V^{mkt} - \Delta\delta S - \sum_k \omega_k \delta U_k^{mkt}$ will grow at risk-free rate too. However, the VV price cannot be a perfect duplicate of V^{mkt}, so the VV hedging strategies ω_k are not actually risk-free.

5.4.2. Black-Scholes Market with Flat Smile. We will now assume that V^{BS} and U_k^{BS} are sufficiently differentiable with respect to the model parameters so we can take Taylor series expansion on them.

The flat vol Black-Scholes model has a single volatility parameter value σ_{ATM}. If the market has a rather flat smile, we can try to expand $V^{mkt} - V^{BS}\big|_{\sigma_{ATM}}$ and $U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}$ around the single σ_{ATM} using Taylor series, where $\delta\sigma$ is the difference between the market Black-Scholes implied vol and σ_{ATM}. By doing so, we actually assume that the systematic error of the model is small and the main error between the model and the market is in the volatility parameter. The Taylor series expansion around the model volatility at the current time leads to

$$\begin{cases} \delta V & \triangleq & V^{mkt} - V^{BS}\big|_{\sigma_{ATM}} \\ & \approx & \dfrac{\partial V^{BS}}{\partial\sigma}\Big|_{\sigma_{ATM}} \delta\sigma + \dfrac{1}{2}\dfrac{\partial^2 V^{BS}}{\partial\sigma^2}\Big|_{\sigma_{ATM}} (\delta\sigma)^2 + O\left((\delta\sigma)^3\right) \\ \delta U_k & \triangleq & U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \\ & \approx & \dfrac{\partial U_k^{BS}}{\partial\sigma}\Big|_{\sigma_{ATM}} \delta\sigma + \dfrac{1}{2}\dfrac{\partial^2 U_k^{BS}}{\partial\sigma^2}\Big|_{\sigma_{ATM}} (\delta\sigma)^2 + O\left((\delta\sigma)^3\right) \end{cases} \tag{5.4.1}$$

where $\delta\sigma = \sigma - \sigma_{ATM}$, σ being the unobservable market Black-Scholes implied volatility for the instrument V.

Note we do not take Itô's expansion here with respect to the current volatility because we are not concerned with the future progression of the volatility, hence no stochasticity is involved.

Substituting (5.2.15) into (5.4.1), we obtain

$$\delta V \approx \sum_k \omega_k \delta U_k + O\left((\delta\sigma)^3\right) \tag{5.4.2}$$

where ω_k satisfy (5.2.15). So we obtain

$$V^{mkt} - \left[V^{BS}\big|_{\sigma_{ATM}} + \sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \right] \approx O\left((\delta\sigma)^3\right) \tag{5.4.3}$$

From (5.4.3), we see that the error between a model price and the sensible mark-to-market price is cubic with respect to volatility using the VV method, given the precondition that the model is a good model describing the market dynamics well but having slightly biased volatility parameter. If the precondition does not hold, the mark-to-market price V cannot be accurately expressed by the Taylor expansion of $V^{BS}\big|_{\sigma_{ATM}}$.

Just from an error analysis point of view, $\left.\dfrac{\partial^2 V^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}} = \sum_k \omega_k \left.\dfrac{\partial^2 U_k^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}}$ is not necessarily needed to ensure a cubic convergence when δS is always zero. However, the condition $\left.\dfrac{\partial^2 V^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}} = \sum_k \omega_k \left.\dfrac{\partial^2 U_k^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}}$ cannot be simply ignored because although the spot risk can be Delta hedged, but the Vega risk brought about by the spot movement can not. That is why the computation of cost of Vega hedging needs to include the effect of $\left.\dfrac{\partial^2 V^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}} = \sum_k \omega_k \left.\dfrac{\partial^2 U_k^{BS}}{\partial\sigma\partial S}\right|_{\sigma_{ATM}}$. We will show that the balance of the Vanna terms is important in a numerical test later.

5.4.3. Market with Non-flat Smile. When the Black-Scholes implied volatility smile is not flat, each strike will have its own implied volatility level. Conceptually, the fact that derivatives for the same underlying but with different strikes can have different implied volatilities seems to be in self-contradiction because the volatility is a property of the underlying. If there is only one underlying, it appears to make more sense to use one volatility for all strikes. Indeed, Derman & Kani [29] observed that prior to the 1987 market crash, there had been no implied smile in the American equity market. Derman & Kani [29] also gave an excellent explanation to the question why varying volatility for strikes makes sense. In the Black-Scholes model, with the constant volatility, the underlying is assumed to follow the lognormal distribution. Traders today do not view the underlying as following the lognormal distribution, which the constant Black-Scholes vol implies. By varying volatility for strikes, traders are essentially changing the underlying distribution

by which they view the market. In today's market, traders tend to view that the out-of-the-money options are more risky than the at-the-money options, hence carrying a risk premium that causes the vol smile to appear. The reasons for a non-lognormal distribution include crashes (downward jumps cause fatter tail on the left side), clustering (One high volatility period is often followed by another high volatility period) etc.

When the market has a smile, it is no longer appropriate to assume that the market follows the Black-Scholes model but with a wrong vol. We will study the error behaviours under different market assumptions later.

5.4.4. The VV Method and the Heston Model. Polishchuk and Carr [82] used the perturbation method to expand the Heston vanilla price with the Black-Scholes price V^{BS}, V_{σ}^{BS}, $V_{\sigma\sigma}^{BS}$, and $V_{\sigma S}^{BS}$.

Assume the Heston model is written as

$$\begin{aligned} \frac{dS}{S} &= (r-q)\,dt + \sqrt{\nu}dW \\ d\nu &= \kappa\left(\theta-\nu\right)dt + \varepsilon\hat{\xi}\sqrt{\nu}dZ \end{aligned}$$

where $dWdZ = \varepsilon\hat{\rho}dt$ and ε is very small. Let $\xi = \varepsilon\hat{\xi}$ and $\rho = \varepsilon\hat{\rho}$.

A Heston call price can be written as

$$\begin{aligned} &V^{Heston} = V^{BS}\left(\hat{\sigma}\right) \\ &+\varepsilon^2\left[-\left(\frac{\alpha\hat{\xi}^2}{\hat{\sigma}^3}-\frac{\gamma}{\hat{\sigma}}\right)V_{\sigma}^{BS}\left(\hat{\sigma}\right)+\alpha\frac{\hat{\xi}^2}{\hat{\sigma}^2}V_{\sigma\sigma}^{BS}\left(\hat{\sigma}\right)+\beta\frac{\hat{\rho}\hat{\xi}}{\hat{\sigma}}SV_{\sigma S}^{BS}\left(\hat{\sigma}\right)\right]+O\left(\varepsilon^3\right) \end{aligned} \tag{5.4.4}$$

where α, β, γ, $\hat{\sigma}$ are functions of κ, θ, ξ, ρ and time to maturity τ. The volatility $\hat{\sigma}$ can be calibrated to the ATM vol σ_{ATM}. While Polishchuk & Carr [82] made the expansion of Heston price around both ξ and ρ, it is worth pointing out that Lewis [65] made a similar expansion around ξ. When ξ and ρ are small, (5.4.4) can be approximately written as

$$\begin{aligned} V^{Heston} &\approx V^{BS}\left(\sigma_{ATM}\right) \\ &+x_1V_{\sigma}^{BS}\left(\sigma_{ATM}\right)+x_2V_{\sigma\sigma}^{BS}\left(\sigma_{ATM}\right)+x_3V_{\sigma S}^{BS}\left(\sigma_{ATM}\right) \end{aligned} \tag{5.4.5}$$

where x_1, x_2, x_3 are functions of the Heston parameters and τ. The three coefficients can be calibrated to three liquid instruments.

A comparison between (5.4.5) and (5.2.31) shows that the Heston model is roughly the same as the VV method for a vanilla option when ξ and ρ are small. (5.4.4) shows that $V^{Heston} \approx V^{VannaVolga} + O\left(\varepsilon^3\right)$. From (5.4.4), the VV method can be viewed as the first

order approximation of the Heston model w.r.t. ξ or ρ. When ξ and ρ are large, the VV price can no longer be viewed as a Heston price due to the existence of the higher order terms in (5.4.4).

Although the results from Polishchuk and Carr [82] show that the Heston vanilla price is about the same as the the VV price for small ξ and ρ, the difference between these two prices depends on the strike. When both the Heston and the VV models are well calibrated to the three liquid strikes, meaning their prices are equal on these strikes, we believe it is reasonable to expect that the price difference on a strike in between the two outer liquid strikes will generally be smaller than that on a strike outside the two outer liquid strikes. The numerical results to be presented later will support this statement. Our numerical results will also show that the small price difference between the VV and Heston prices is not limited to small ξ and ρ.

Polishchuk and Carr [82] suggest that for exotic options, the above perturbation results will not hold, and hence the above correspondence between the Heston and VV prices at small ξ and ρ does not apply.

5.4.5. The VV Method and the Gram-Charlier Series.

5.4.5.1. *Option Pricing using the Gram-Charlier Series.* The Black-Scholes model assumes the normal asset return probability distribution. If the probability distribution of an asset return is nonnormal, it is possible to express it with an expansion series based on the normal distribution and the moments of the nonnormal distribution. This expansion series is commonly known as the Gram-Charlier series in the literature [19, 31, 17, 28]. Following [23], the Gram-Charlier series in a general setting is expressed by

$$P(z) \approx P^m(z) = \sum_{n=0}^{m} c_n H_n(z)\, \phi(z) \tag{5.4.6}$$

where $P(z)$ is the nonnormal probability distribution, $P^m(z)$ is the approximation of $P(z)$ using the first m terms of the series, $\phi(z)$ is the standard normal distribution with zero mean and unit variance, and $H_n(z)$ are the Hermite polynomials

$$H_n(z) = \sum_{k=0}^{n/2} \frac{(-1)^k\, n!}{2^k k!\,(n-k)!} z^{n-2k}, \tag{5.4.7}$$

and the coefficients c_n are determined by

$$c_n = \frac{1}{n!} \int_{-\infty}^{\infty} H_n(z) P(z) dz \tag{5.4.8}$$

Jarrow & Rudd [50], Corrado & Su [24, 26, 25], Knight and Satchell [59] applied the Gram–Charlier series to the Black–Scholes formula for vanilla options in order to model the market skew and kurtosis in the distribution of stock returns. Madan & Milne [68], Longstaff [67], Abken et al [1], Brenner & Eom [12], Potters et al [84], Backus et al [4] have applied the Gram–Charlier series to the studies of financial economics. Marco Airoldi [3] applied the Gram–Charlier series to Asian options with discrete fixings, reverse cliquet options, and barrier options with discrete monitoring dates.

The Gram–Charlier series is usually truncated to the fourth moment. Corrado & Su [24] suggest that adding more higher order moments to the series often leads to unstable parameter estimates instead of better accuracy; for practical applications, it is recommended to use up to the fourth moment. Because of the truncation, the Gram-Charlier series with limited moment terms may not always be positive [28]; hence strictly speaking, it is not a valid probability density distribution but an approximation to a probability density distribution. A nonnormal probability distribution P can thus be expressed by the Gram-Charlier probability distribution $P^{GC}(z)$ [26]

$$P(z) \approx P^{GC}(z) = \phi(z) \left[1 + \frac{\mu_3}{3!} \left(z^3 - 3z \right) + \frac{\mu_4 - 3}{4!} \left(z^4 - 6z^2 + 3 \right) \right] \tag{5.4.9}$$

where μ_3 and μ_4 are the normalised skew and kurtosis of $P(z)$, and z is defined by

$$z = \frac{\ln(S_T/S_0) - \left(r - \sigma_{GC}^2/2 \right) T}{\sigma_{GC}\sqrt{T}}, \tag{5.4.10}$$

and S_T is the underlying price at the maturity, S_0 is the current underlying price, r is the interest rate, σ_{GC} is a volatility parameter to be calibrated from the market, and T is the time to maturity.

Corrado & Su [26] gives the call option price by integrating the payoff function

$$\begin{aligned} C &= e^{-rT} \int_K^{\infty} (S_T - K) P(z(S_T)) dz(S_T) \\ &= C_{BS} + \mu_3 Q_3 + (\mu_4 - 3) Q_4 \end{aligned} \tag{5.4.11}$$

where C_{BS} is the Black-Scholes price (3.2.3),

$$\begin{aligned} Q_3 &= \frac{1}{3!}S_0\sigma_{GC}\sqrt{T}\left(\left(2\sigma_{GC}\sqrt{T}-d\right)\phi(d)-\sigma_{GC}^2T\mathcal{N}(d)\right), \\ Q_4 &= \frac{1}{4!}S_0\sigma_{GC}\sqrt{T}\left(\left(d^2-1-3\sigma_{GC}\sqrt{T}\left(d-\sigma_{GC}\sqrt{T}\right)\right)\phi(d)+\sigma_{GC}^3T^{3/2}\mathcal{N}(d)\right), \\ d &= \frac{\log(S_0/K)+(r-\frac{1}{2}\sigma_{GC}^2)T)}{\sigma_{GC}\sqrt{T}}, \end{aligned}$$

and $\phi(\,)$ is the standard normal distribution, $\mathcal{N}(\,)$ is the cumulative normal function (3.2.4).

Since the model parameters σ_{GC}, μ_3 and μ_4 are to be calibrated to the market prices, the Gram-Charlier model involve potentially time-consuming and inaccurate nonlinear optimisation. Compared to the Gram-Charlier model, the VV method has a clear advantage in not requiring calibration.

5.4.5.2. *Probability Density Function of the VV Method.* Breeden and Litzenberger [11] suggest that the state price probability density function can be obtained by taking the second derivative of the call option price with respect to the strike. Castagna and Mercurio [15] give the probability density function at the maturity based on the VV price

$$\begin{aligned} p^{VV}(K) &= e^{rT}\frac{\partial^2C^{VV}}{\partial K^2}(K)=e^{rT}\frac{\partial^2C^{BS}}{\partial K^2}(K) \\ &\quad +e^{rT}\frac{\partial^2\omega_1}{\partial K^2}\left(U_1^{mkt}-U_1^{BS}\right)+e^{rT}\frac{\partial^2\omega_3}{\partial K^2}\left(U_3^{mkt}-U_3^{BS}\right) \end{aligned} \tag{5.4.12}$$

Notice the ATM price difference $U_{ATM}^{mkt}-U_{ATM}^{BS}=0$ because the Black-Scholes implied vol is calibrated to the ATM vol, and $e^{rT}\frac{\partial^2C^{BS}}{\partial K^2}(K)$ is the the normal distribution $g(z)$. We will use the following approximate equalities for the vanillas:

$$\begin{aligned} U_1^{mkt}-U_1^{BS} &\approx \left.\frac{\partial\left(U_1^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}}(\sigma_1-\sigma_{ATM}) \\ U_3^{mkt}-U_3^{BS} &\approx \left.\frac{\partial\left(U_3^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}}(\sigma_3-\sigma_{ATM}) \end{aligned} \tag{5.4.13}$$

where σ_1 is the Black-Scholes implied vol of U_1^{mkt} and σ_3 is the Black-Scholes implied vol of U_3^{mkt}.

Combining (5.4.13) and (5.4.12) gives

$$p^{VV}(K)\approx\phi(z(K))+A(K)\left(RR^{mkt}\right)+B(K)\left(BF^{mkt}\right) \tag{5.4.14}$$

where the risk reversal $RR^{mkt} = \sigma_3 - \sigma_1$, the butterfly $BF^{mkt} = \frac{\sigma_1 + \sigma_3}{2} - \sigma_2$, and

$$A(K) = \frac{1}{2}\left(e^{rT}\frac{\partial^2 \omega_3}{\partial K^2}\left.\frac{\partial\left(U_3^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}} - e^{rT}\frac{\partial^2 \omega_1}{\partial K^2}\left.\frac{\partial\left(U_1^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}}\right),$$
$$B(K) = e^{rT}\frac{\partial^2 \omega_1}{\partial K^2}\left.\frac{\partial\left(U_1^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}} + e^{rT}\frac{\partial^2 \omega_3}{\partial K^2}\left.\frac{\partial\left(U_3^{BS}\right)}{\partial\sigma}\right|_{\sigma_{ATM}}.$$

$A(K)$ and $B(K)$ are function of K because ω_k are functions of K.

Combining (5.4.14) with (5.1.12) gives

$$p^{VV}(K) \approx \phi(z(K)) + \frac{\mu_3}{3!}\widetilde{A}(K) + \left(\frac{\mu_4 - 3}{4!}\right)\widetilde{B}(K) \tag{5.4.15}$$

$$\begin{aligned}\widetilde{A}(K) &\approx A(K)(k_c - k_p)\sigma_{ATM}\\ \widetilde{B}(K) &\approx B(K)\left(\frac{\left(k_c^2 + k_p^2\right)}{2} - k_{ATM}^2\right)\sigma_{ATM}\end{aligned}$$

$\widetilde{A}$ and $\widetilde{B}$ are not the same as the Hermite polynomials H_3 and H_4 described by (5.4.7) so the probability density functions p^{VV} and p^{GC} are not exactly equal.

5.5. Performance of the VV Method for Vanillas

We will study the performance of the VV method under different market assumptions. First, we will assume that the Black-Scholes model follows the market but the Black-Scholes implied vol used for the model is wrongly estimated. Next, we will test how the accuracy will be under the Heston, CEV, SABR and Merton Jump models as the market.

5.5.1. Model Follows Market but with Wrong Parameter Estimation. In this section, we will present results for the cases where the Black-Scholes model resembles the market but with wrong volatility estimation. The market has the market volatility but the model adopts a volatility slightly different from the market vol. The case of a model potentially using a model parameter value different from the true market value is quite common in practice.

The market is assumed to be a vanilla call option with the market Black-Scholes implied vol, and the model is the same call option but with a mispriced Black-Scholes implied vol. This instrument and the hedging vanillas have the following parameters: the maturity $T = 1$, the spot $S = 5$, the domestic rate $r_d = 0.03$, the foreign rate $r_f = 0.02$, the market Black-Scholes implied vol $\sigma_{mkt} = 0.25$, the 25$\triangle$ put strike $K_{25p} = 4.42$, at-the-money

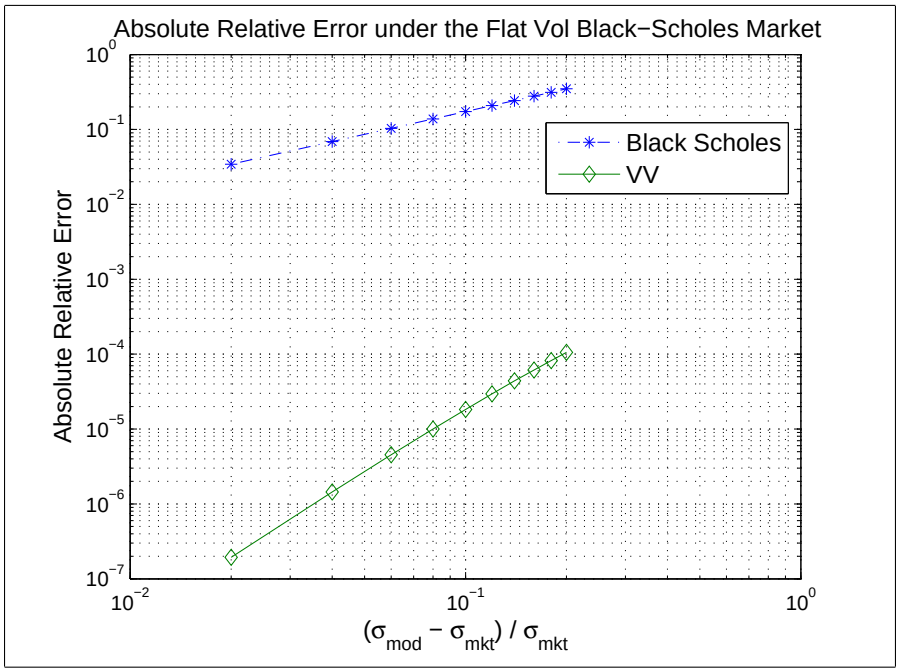

FIGURE 5.5.1. Cubic Convergence of the VV Price Error in a Flat Vol Black-Scholes Market. (See text for parameter values)

strike $K_{ATM} = 5.21$, 25$\triangle$ call strike $K_{25c} = 6.14$, the strike of the instrument to be priced $K = K_{ATM} + 0.5\,(K_{25c} - K_{ATM})$. Figure (5.5.1) compares the convergence behaviours of the Black-Scholes price and the VV price. With log scales, the rate of convergence is clearer. The model Black-Scholes implied vol σ_{mod} shifts away from the market Black-Scholes implied vol σ_{mkt} by some percentage that is expressed by the $(\sigma_{mod} - \sigma_{mkt})/\sigma_{mkt}$ axis.

The resulting absolute relative error of the Black-Scholes price is $|V(\sigma_{mkt}) - V(\sigma_{mod})|/V(\sigma_{mkt})$, which in this plot demonstrates a first-order convergence. By comparison, the absolute relative error of the VV price $\left|V(\sigma_{mkt}) - V^{VV}\right|/V(\sigma_{mkt})$ achieves the far better third-order convergence. Not only is the third-order convergence result outstanding, the absolute value of the VV price error drastically drops to a level that is negligible whereas the absolute value of the Black-Scholes price errors are still in the range of dozens of basis points in this case. This means that the VV price performs much better than the model price alone in absorbing the error caused by the model parameter (Black-Scholes implied volatility) mispricing. The VV price is a rather stable price that is capable of retrieving the true price even if the model parameter is slightly wrong. The amazing characteristic of the VV price comes from the market information the vanillas on the vol pillars provide.

5.5.2. Heston as the Market - Model Dynamics Do not Follow the Market. We have illustrated the situation where the market is more or less the Black-Scholes world but the Black-Scholes model is used with imperfect model vol parameter that is different from its market counterpart. Now we will present comparisons made between a market and model that are not so similar in their dynamics.

In the Black-Scholes world, the concept of volatility is expressed by a single parameter, the Black-Scholes implied volatility. The Heston model does not use the Black-Scholes implied vol as its sole volatility parameter. Rather it relies on several parameters to describe the volatility of the market. The Heston model is described by

$$\begin{aligned} dS &= (r_d - r_f)\, S dt + S\sqrt{\nu} dW \\ d\nu &= \kappa\,(\theta - \nu)\, dt + \xi\sqrt{\nu} dZ \end{aligned} \tag{5.5.1}$$

where the correlation of dW_t and dZ_t is ρ. Elder [32] in a study of hedging strategies used the Heston model as the assumed market. To understand better how the Vanna-Volga price is different from that of the stochastic model, we will follow Elder [32] to use the Heston model as the market truth.

Figure (5.5.2) shows the Black-Scholes implied vol comparison of the Heston, Black-Scholes and VV models. The parameters for the Heston model and the option are $\kappa = 1.1$, $\theta = 0.09$ $\xi = 0.27$, $\rho = -0.7$, $\nu_0 = 0.09$, $\tau = 0.6$, $S = 5$, $r_d = 0.03$ and $r_f = 0.02$. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The Black-Scholes price is calibrated at the strike K_2 against the Heston price on K_2. So the ATM Black-Scholes implied vol for the VV method is backed out by the Heston price on K_2. The weights ω_k in the VV price formula are computed using the Heston prices on the liquid strikes, hence the VV and Heston prices exactly agree on the strikes K_1, K_2, and K_3. Figure (5.5.3) gives the relative error between the Heston and VV prices. The relative error is defined by

$$\text{Relative Error} = \frac{\text{Error}}{V^{market}}. \tag{5.5.2}$$

where $\text{Error} = V^{market} - V^{estimate}$.

In this test, $V^{market} = V^{Heston}$ and $V^{estimate} = V^{VV}$. Figures (5.5.2) and (5.5.3) show that between K_1 ($K_1/S = 0.86$) and K_3 ($K_3/S = 1.14$) the price difference between the Heston model and the VV method is small (the relative error is less than 0.05%). Outside of the range of interpolation between K_1 and K_3, the price difference increases quickly

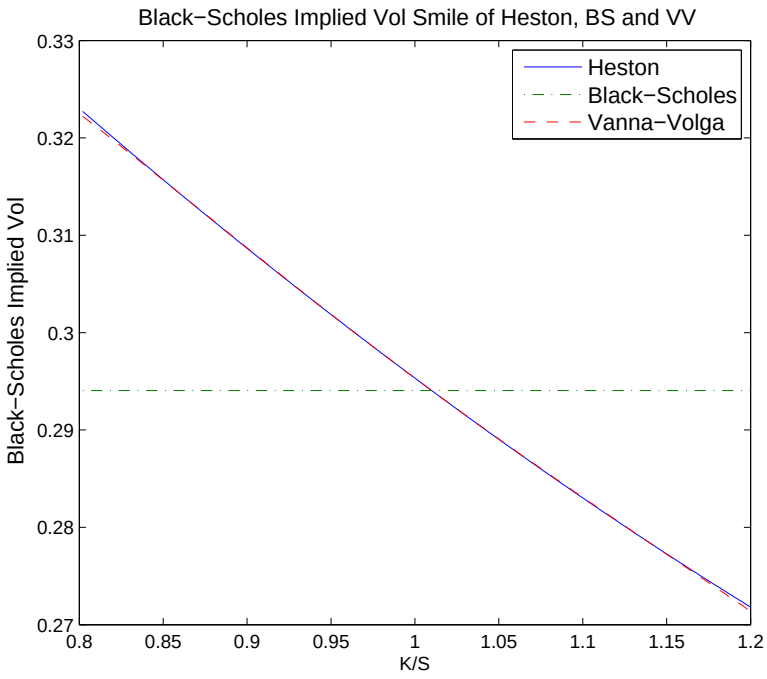

FIGURE 5.5.2. Black-Scholes Implied Vol Smile of Heston, Black-Scholes and VV

but a range of extrapolation still allows small price difference between the Heston and VV models. Overall, the fitting of the VV price to the Heston price is excellent in the range of $0.8 < K/S < 1.2$. The Black-Scholes model calibrated to the ATM vol displays increasing differences with the Heston price as K is further away from the strike K_2.

Figures (5.5.4), (5.5.5), and (5.5.6) display the behaviours of the relative error defined in (5.5.2) when the parameters are varied. The option in these comparisons is a European call at $K = 0.5\,(K_2 + K_3) = 5.35$. We use this strike for more testing because (5.5.3) shows that the price difference between the Heston and VV models at K is relatively bigger compared to other strikes between K_1 and K_3. The top plot of Figure (5.5.4) is the same as the bottom plot except that the bottom plot shows the data on $-1 < \rho < 0$ and $0 < \xi < 0.5$ whereas the top plot shows the data on $-1 < \rho < -1$ and $0 < \xi < 1$. Of the two plots of Figure (5.5.4), the ranges of ρ and ξ in the bottom plot are usually the realistic choices for the Heston model in the market. For the normal ranges of parameters, Figures (5.5.4), (5.5.5), and (5.5.6) show that the relative errors between the Heston and VV prices are small, usually in the order of 10^{-3}. This suggests that the VV price is an excellent alternative to the Heston model, at least for vanilla options.

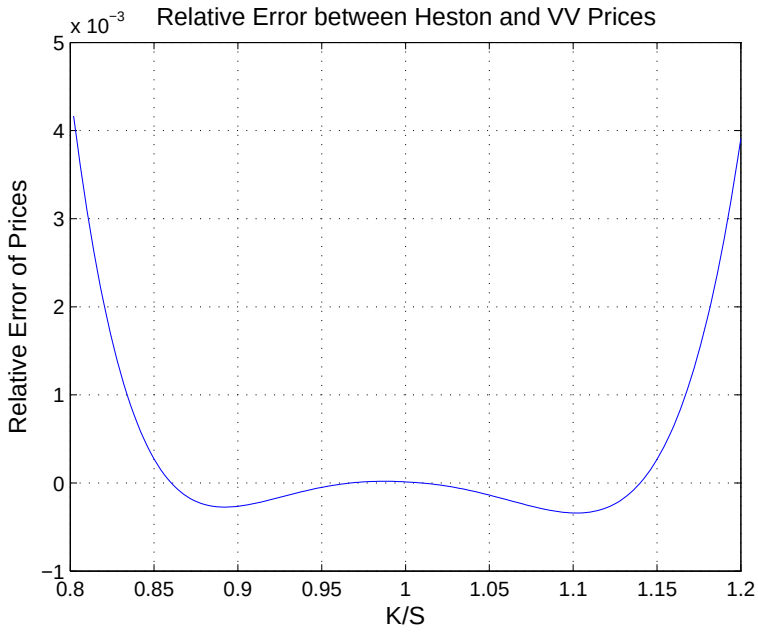

FIGURE 5.5.3. Relative Error between Heston and VV Prices

5.5.2.1. *Comparison with the Cases of One Hedging Vanilla and Two Hedging Vanillas.* Echoing the earlier analysis made on the relationship between $(d\sigma)^2$, $(d\sigma dS)$ and dt and the argument whether less than three hedging vanillas would be good enough, we will study the effect of reducing the number of vanillas used in the VV formula on the Heston-VV price difference. The parameters are the same as the previous tests.

With the option with strike K_2 as the only Hedging vanilla, the VV price formula will be

$$V^{VV1} = V^{BS} + \omega \left(U_2^{mkt} - U_2^{BS} \right), \quad \text{where} \quad \left. \frac{\partial V^{BS}}{\partial \sigma} \right|_{\sigma_{ATM}} = \omega \left. \frac{\partial U_2^{BS}}{\partial \sigma} \right|_{\sigma_{ATM}}$$

With both K_2 and K_3 as the Hedging vanillas, the VV price formula will be

$$V^{VV2} = V^{BS} + \omega_2 \left(U_2^{mkt} - U_2^{BS} \right) + \omega_3 \left(U_3^{mkt} - U_3^{BS} \right)$$

where

$$\left(\begin{array}{c} \dfrac{\partial V^{BS}}{\partial \sigma} \\ \dfrac{\partial^2 V^{BS}}{\partial \sigma^2} \end{array} \right) \Bigg|_{\sigma_{ATM}} = \left(\begin{array}{cc} \dfrac{\partial U_2^{BS}}{\partial \sigma} & \dfrac{\partial U_3^{BS}}{\partial \sigma} \\ \dfrac{\partial^2 U_2^{BS}}{\partial \sigma^2} & \dfrac{\partial^2 U_3^{BS}}{\partial \sigma^2} \end{array} \right) \Bigg|_{\sigma_{ATM}} \left(\begin{array}{c} \omega_2 \\ \omega_3 \end{array} \right)$$

Figure (5.5.7) shows the difference between the Heston and VV prices using only one or two hedging vanillas. The model and option parameters are the same as the case of the three hedging vanillas in the assumed Heston market. The left plots are shown with Black-Scholes implied vol smiles and the right plots are shown with relative errors (defined

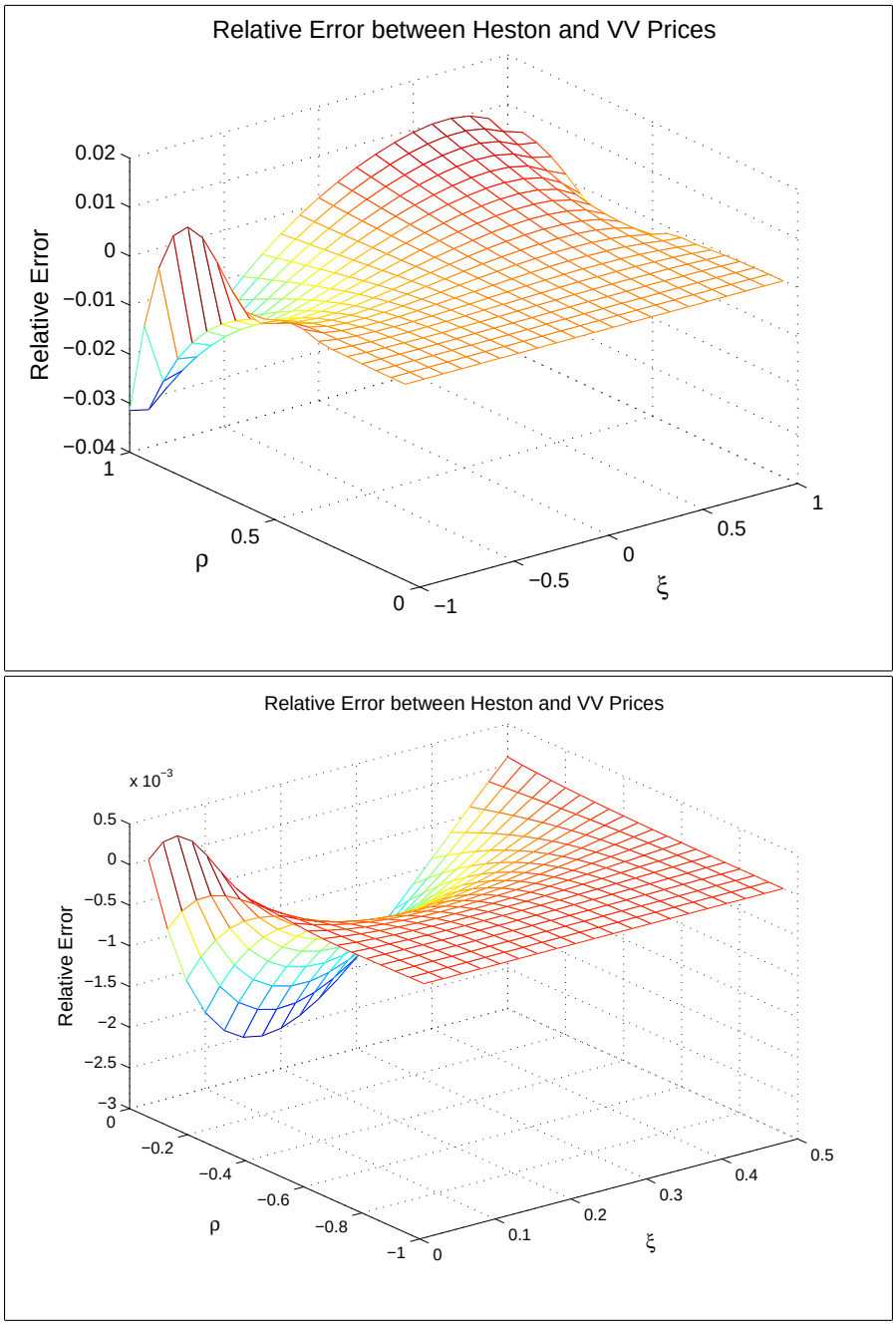

FIGURE 5.5.4. Relative Error between Heston and VV vs ρ and ξ

in (5.5.2)) of the option prices . The top plots are made with one hedging vanilla and the bottom plots are made with two hedging vanillas. With one hedging vanilla at K_2, the VV formula can only perfectly fit the price at K_2. With two hedging vanillas at K_2 and K_3, the VV formula can perfectly fit the prices at K_2 and K_3. The relative error increases quickly when K is away from K_2 with one hedging vanilla. With two hedging vanillas, the relative error is held relatively small between the strikes of these two hedging vanillas. But

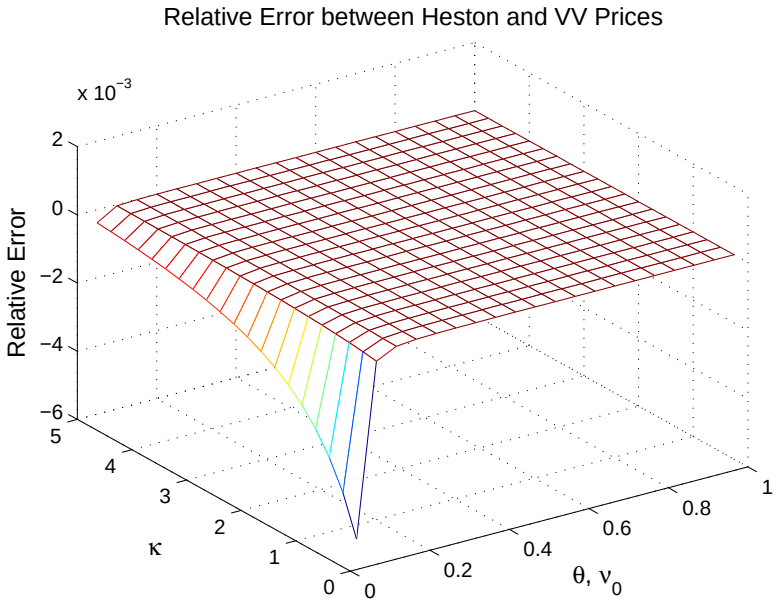

FIGURE 5.5.5. Relative Error between Heston and VV vs θ, ν_0 and κ

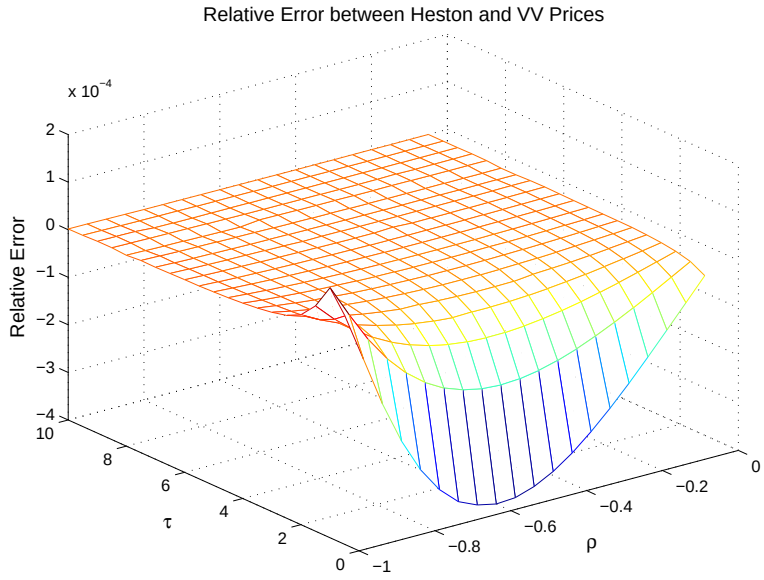

FIGURE 5.5.6. Relative Error between Heston and VV vs τ and ρ

a comparison between (5.5.7) and (5.5.3) reveals that the difference using three hedging vanillas is much smaller than the two-hedging-vanillas case between K_2 and K_3.

Figure (5.5.8) shows the plots of relatives errors vs ρ and ξ. The option used is a European call at $K = 0.5\,(K_2 + K_3)$. Although the relative error of the case of two hedging vanillas is smaller than that of the case of one hedging vanilla, a comparison with

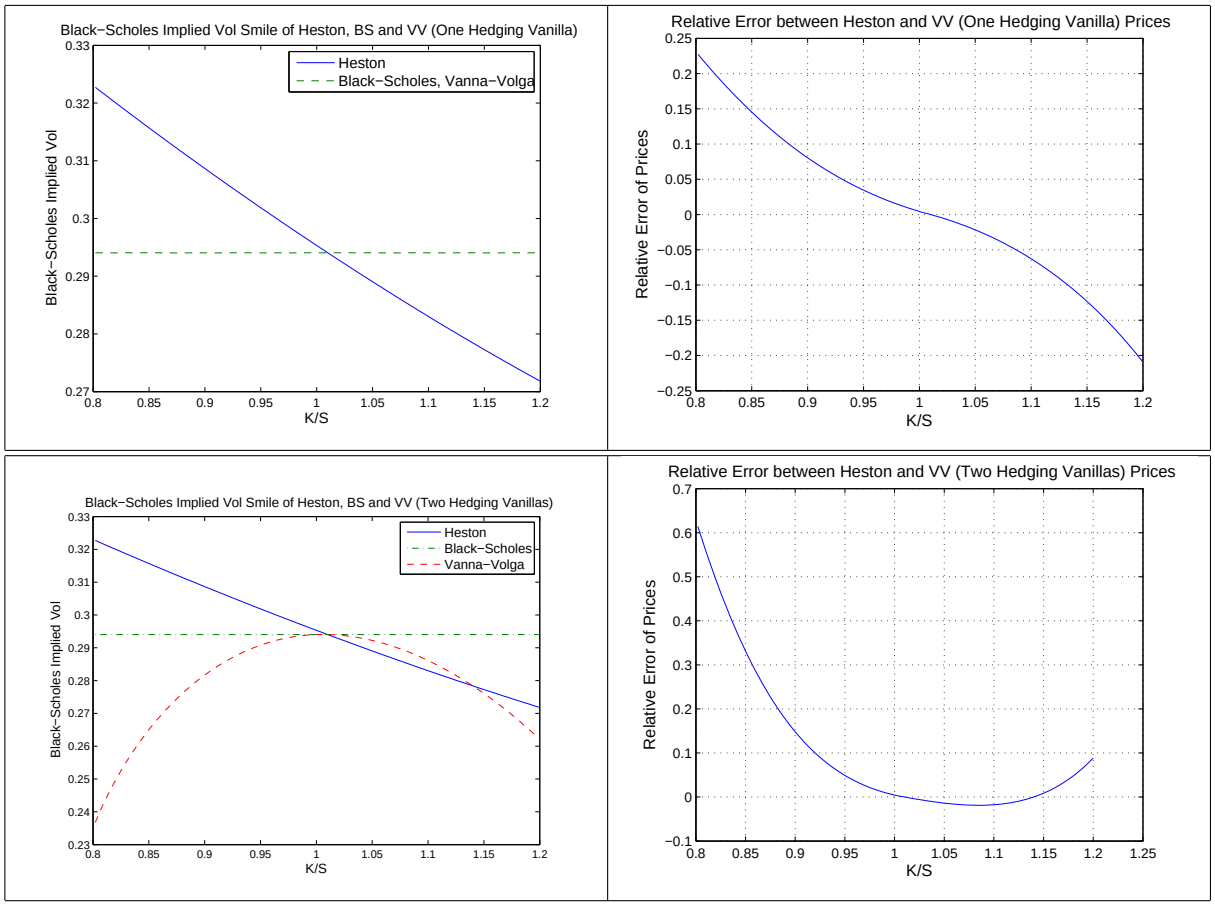

FIGURE 5.5.7. Heston, Black-Scholes and VV with One/Two Hedging Vanillas

the case of the three hedging vanillas in Figure (5.5.4) reveals that the relative error of the three hedging vanillas is much smaller than either of them.

5.5.2.2. *ω_k Computed by Vega, Volga, and $\frac{\partial Volga}{\partial \sigma}$ Using Three Liquid Vanillas.* In the VV method, ω_k is computed by balancing the Vega, Volga and Vanna of V^{BS} and the ω_k weighted sum of Vega Volga, and Vanna of U_k^{BS}. From an error analysis view, the choices of Vega, Volga and Vanna are not arbitrary. For a flat vol market, as shown earlier, the choices of Vega, Volga and Vanna cancel out the first and second order error terms about the Black-Scholes implied vol.

We will explore how the error behaviour changes when this Vega, Volga, Vanna combination is altered. In the combination of Vega, Volga and Vanna, we will replace Vanna with $\frac{\partial Volga}{\partial \sigma}$, where σ is the Black-Scholes implied vol. As analysed earlier, using $\frac{\partial Volga}{\partial \sigma}$ instead of Vanna will still allow the VV method to fit the market prices on the liquid strikes. From an error analysis view, using a higher order sensitivity but leaving a lower

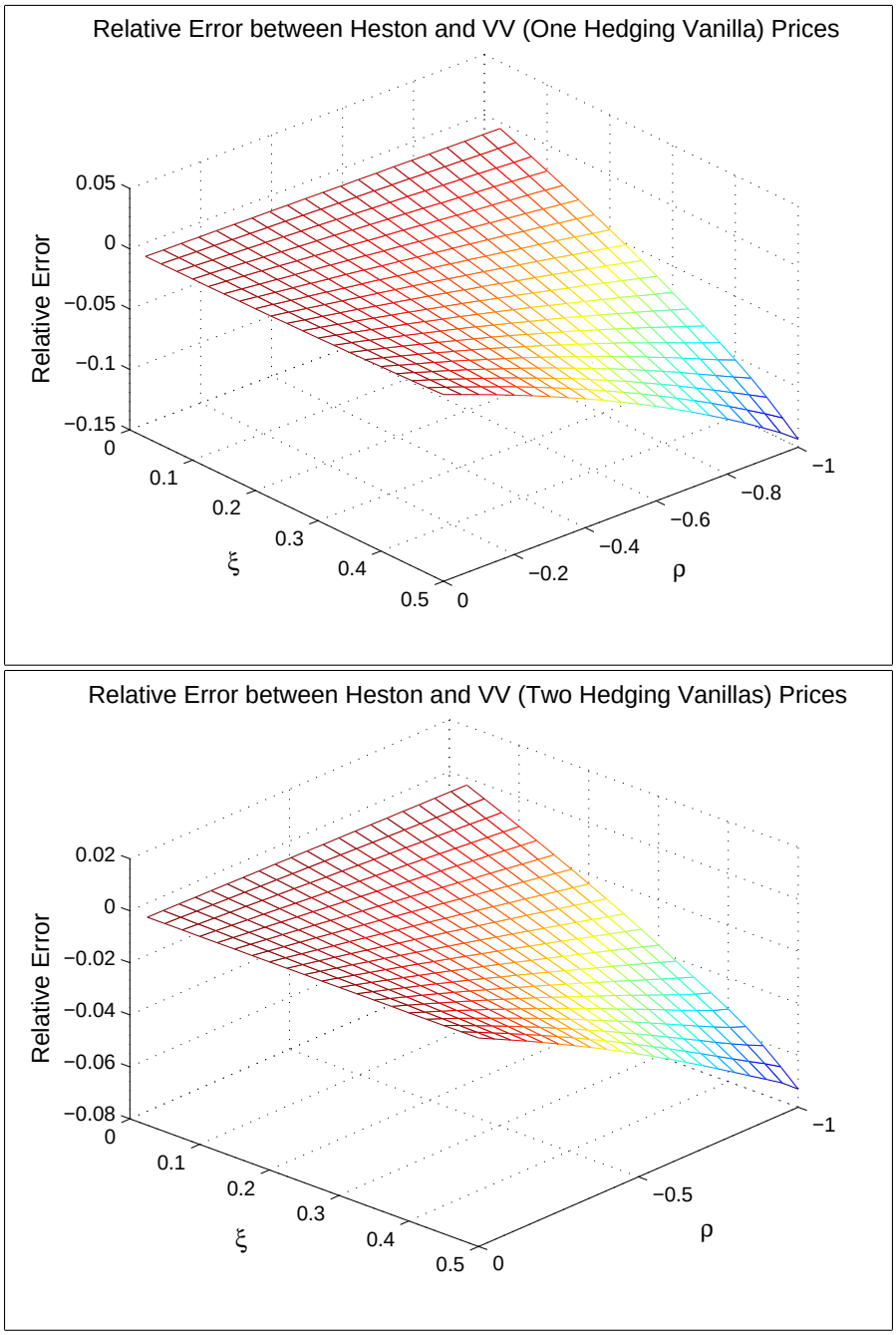

FIGURE 5.5.8. Relative Error between Heston and VV vs ρ and ξ with One/Two Hedging Vanillas

order sensitivity unbalanced will likely make the error not on the liquid strikes become bigger.

We will assume the Heston model follows the market. The parameters used in Figure (5.5.9) are the same as those used in Figures (5.5.2) and (5.5.3). The only difference is that in Figure (5.5.9), ω_k are computed with the sensitivities Vega, Volga, and $\frac{\partial Volga}{\partial \sigma}$ instead of Vega, Volga, and Vanna. The left plot of Figure (5.5.9) shows the comparison

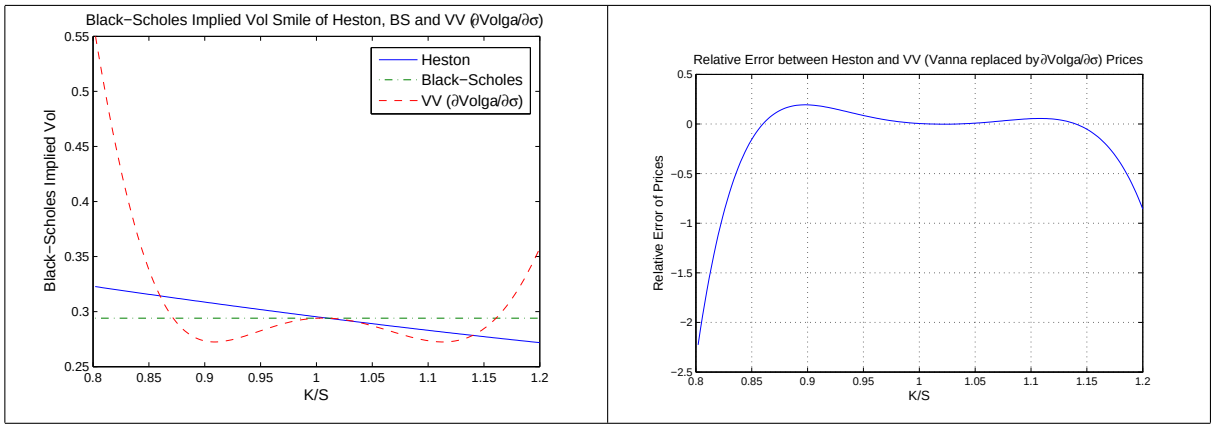

FIGURE 5.5.9. Heston, Black-Scholes and VV (Vanna Replaced by $\frac{\partial Volga}{\partial \sigma}$)

of implied vol smiles and the right plot shows the relative error between the Heston and VV prices.

As analysed earlier, the market prices of the three liquid strikes are perfectly produced by the VV formula using $\frac{\partial Volga}{\partial \sigma}$ in Figure (5.5.9). But the error elsewhere is much bigger than the usual VV method using Vanna in computation of ω_k. Therefore, the combination choice of sensitivities is not arbitrary.

5.5.3. CEV as the Market. The Constant Elasticity of Variance (CEV) model [48] assumes that the underlying satisfies

$$dS = (r_d - r_f)\, S dt + \sigma S^{\alpha} dW \tag{5.5.3}$$

where dW is a Wiener process, σ is the volatility parameter of the CEV model, and the constant $\alpha > 0$. The CEV model volatility of the underlying return is $\sigma S^{\alpha-1}$. When $\alpha = 1$, the CEV model is just the Black-Scholes model; when $0 < \alpha < 1$, the CEV model volatility of the underlying return increases as S decreases, giving the probability distribution a heavy left tail and a less heavy right tail; when $\alpha > 1$, the CEV model volatility of the underlying return increases as S increases, giving the probability distribution a heavy right tail and a less heavy left tail. While the heavy left tail distribution may be more common, the heavy right tail is sometimes seen in options on futures [48]. The CEV model has the following

formulas [48] for European call and put

$$V_{call} = \begin{cases} Se^{-r_f\tau}\left(1-\chi^2\left(a,\, b+2\,,c\right)\right) - Ke^{-r_d\tau}\chi^2\left(c,\, b,\, a\right), & 0<\alpha<1 \\ Se^{-r_f\tau}\left(1-\chi^2\left(c,\, -b\,,a\right)\right) - Ke^{-r_d\tau}\chi^2\left(a,\, 2-b,\, c\right), & \alpha>1 \end{cases}$$

$$V_{put} = \begin{cases} Ke^{-r_d\tau}\left(1-\chi^2\left(c,\, b\,,a\right)\right) - Se^{-r_f\tau}\chi^2\left(a,\, b+2,\, c\right), & 0<\alpha<1 \\ Ke^{-r_d\tau}\left(1-\chi^2\left(a,\, 2-b\,,c\right)\right) - Se^{-r_f\tau}\chi^2\left(c,\, -b,\, a\right), & \alpha>1 \end{cases}$$

where

$$a=\frac{K^{2(1-\alpha)}}{(1-\alpha)^2\sigma^2\tau},\ b=\frac{1}{(1-\alpha)},\ c=\frac{\left(Se^{(r_d-r_f)\tau}\right)^{2(1-\alpha)}}{(1-\alpha)^2\sigma^2\tau}$$

and χ^2 is the cumulative non-central χ^2 function. In Matlab, this function is ncx2cdf.

The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The parameters for the CEV model and the option are $\alpha = 0.6$, $\sigma = 0.25$, $\tau = 1$, $S = 5$, $r_d = 0.03$, and $r_f = 0.02$. Figure (5.5.10) shows the Black Scholes implied volatilities of the CEV and VV prices for different strikes. Figure (5.5.11) shows the relative error between the CEV and VV prices for different strikes; put prices are used before K_2 and call prices are used after K_2 to compute the relative error. For the range between K_1 and K_3, the relative error is less than 0.05%.

Figure (5.5.12) displays the relative error between the CEV an VV prices for a European put at $K = (K_1 + K_2)/2 = 4.675$ with respect to changes in σ and α. When both $0 < \sigma < 0.1$ and $\alpha < 0.8$ are satisfied, the relative error will increase as both σ and α decrease. But for the most part of $0 < \sigma < 1$ and $0.5 < \alpha < 2$, the relative error is less than 0.05%.

5.5.4. SABR as the Market. Hagan et al [40] proposed the SABR (meaning "Stochastic Alpha, Beta, Rho") model, which is as follows:

$$\begin{aligned} dF &= \alpha F^{\beta} dW_1 \\ d\alpha &= \nu\alpha dW_2 \end{aligned} \tag{5.5.4}$$

where $dW_1 dW_2 = \rho dt$, $-1 \leqslant \rho \leqslant 1$, F is the forward price of the underlying, α is the stochastic volatility, ν is the vol of vol ($\nu \geqslant 0$), and $0 \leqslant \beta \leqslant 1$. The SABR model is like a CEV model with a stochastic volatility.

Castagna and Mercurio [16] gave results on the excellent agreement between the Black-Scholes implied vols of the SABR and VV prices. We will show the relative error between

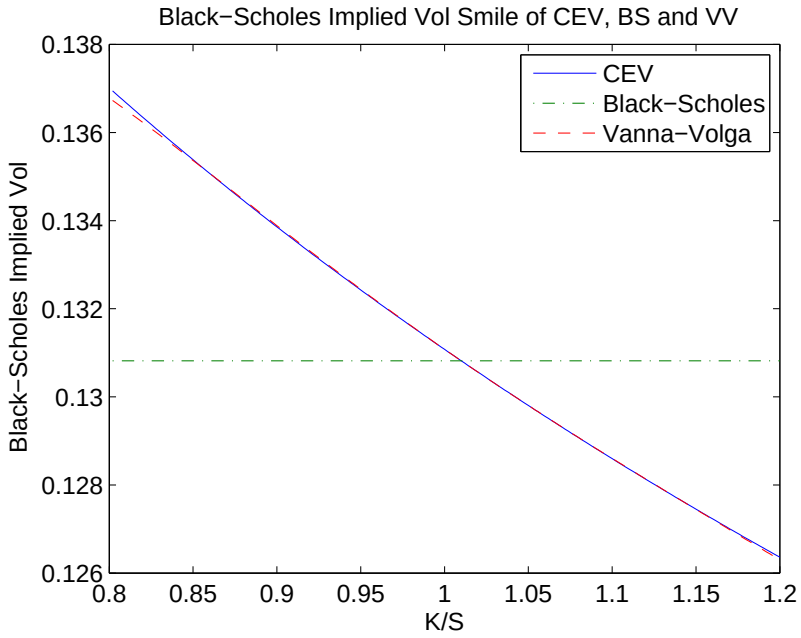

FIGURE 5.5.10. Black Scholes Implied Vols of the CEV and VV Models

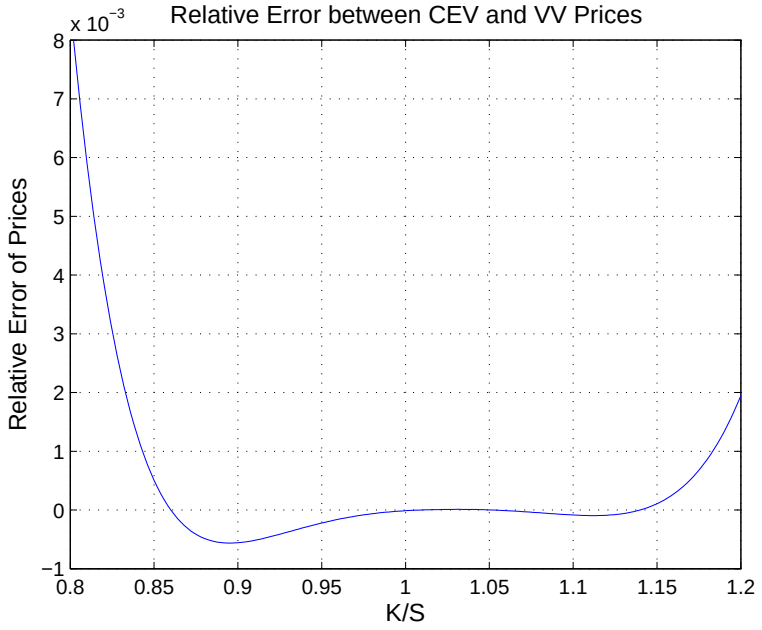

FIGURE 5.5.11. Relative Error between CEV and VV Prices

the SABR and VV prices in Figure (5.5.14). The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The parameters for the SABR model and the option are $\beta = 0.5$, $\rho = -0.5$, $\nu = 0.3$, $\tau = 1$, $S = 5$, $r_d = 0.03$, $r_f = 0$, the current vol $\alpha_0 = 0.4$. Figure (5.5.14) shows that within the range of the three strikes, the prices are very close (the relative error smaller than 0.5%).

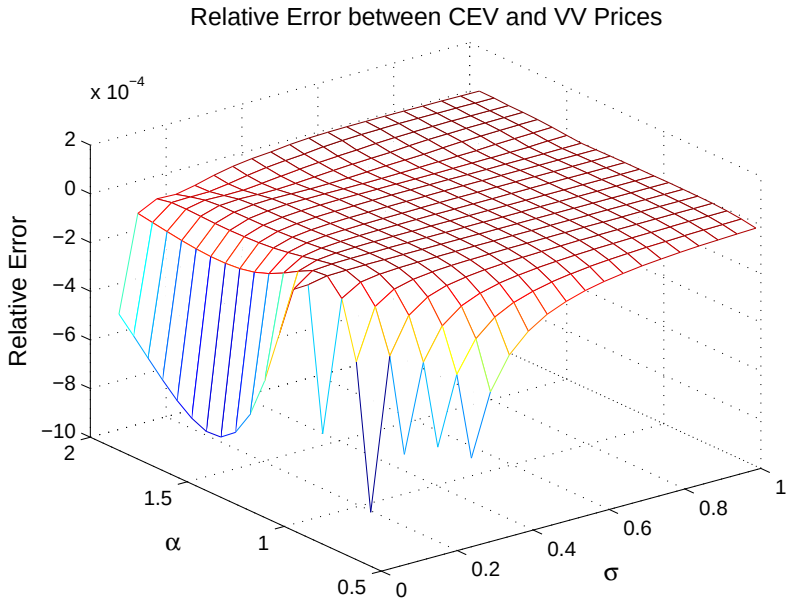

FIGURE 5.5.12. Relative Error between CEV and VV Prices

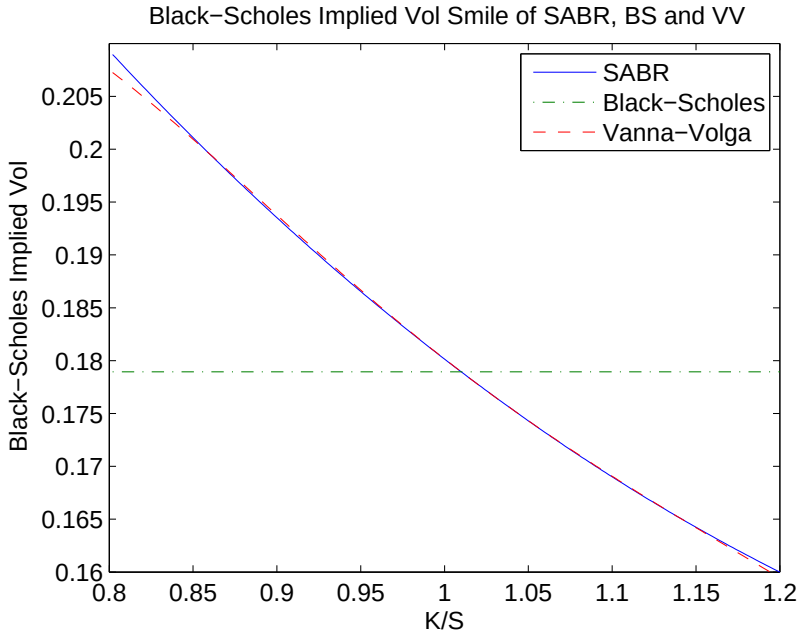

FIGURE 5.5.13. Black Scholes Implied Vols of SABR and VV

We will further show how the agreement between the SABR and VV prices changes with the model parameters β, α_0, ρ, ν. Assuming the above parameters, Figure (5.5.15) and Figure (5.5.16) show the relative errors between the SABR and VV prices for a European put option at $K = (K_1 + K_2)/2 = 4.675$ with respect to changes in model parameters. Because the negative ρ is arguably more common, the right plot of Figure (5.5.16) which

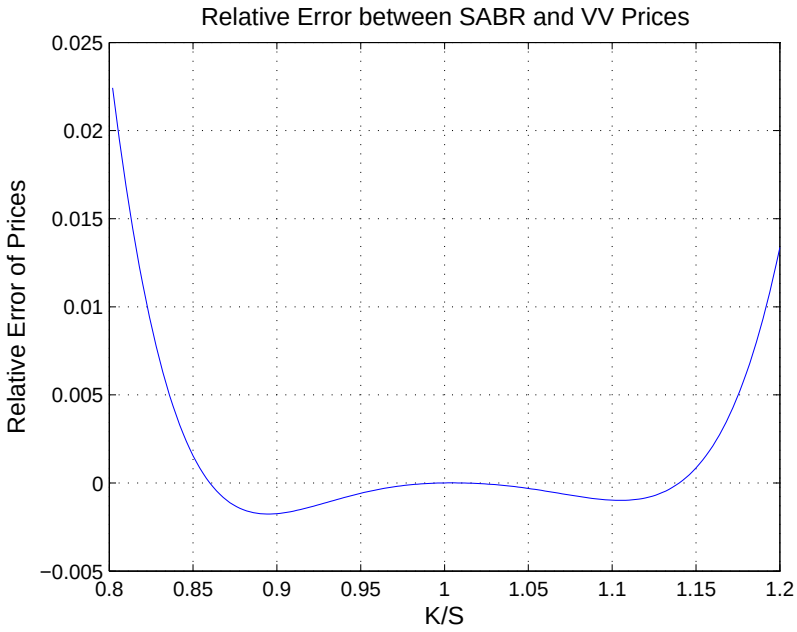

FIGURE 5.5.14. Relative Error between SABR and VV prices

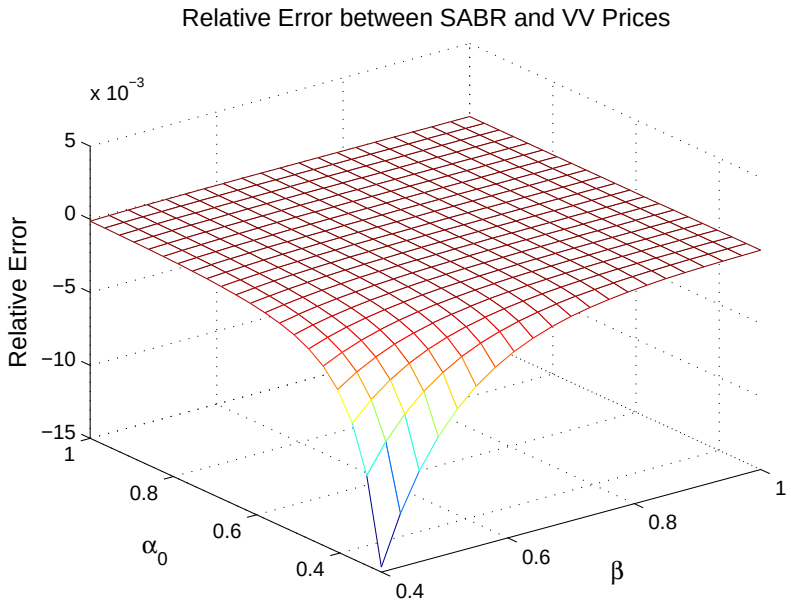

FIGURE 5.5.15. Relative Error between SABR and VV prices

is a part of the left plot is shown as a separate plot. For most of the ranges shown in these plots, the relative error is in the order of 10^{-3}.

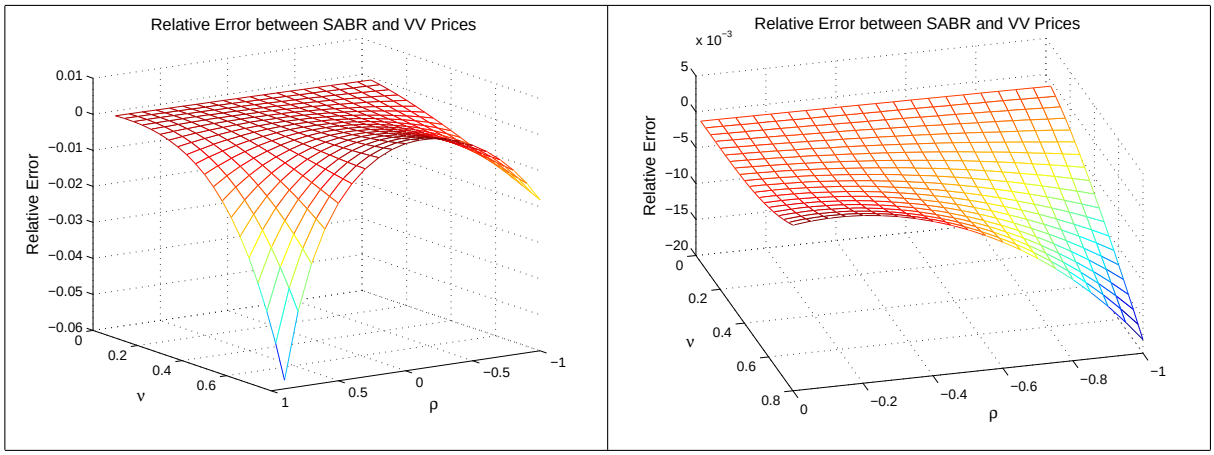

FIGURE 5.5.16. Relative Error between SABR and VV prices

5.5.5. Merton Jump Diffusion[1] as the Market. Merton [71] developed the following jump diffusion model

$$dS = (r_d - r_f)\, Sdt + \sigma SdW + (J-1)\, SdN \tag{5.5.5}$$

where σ is the volatility of the geometric Brownian motion, dW is a Wiener process, dN is the number of jumps that happen in dt, N follows the Poisson process $P(N=n) = (\lambda\tau)^n e^{-\lambda\tau}/n!$, λ is the average number of jumps per year. The probability of a jump in dt is λdt. The processes dW and dN are independent of each other.

J usually is assumed to follow the lognormal distribution with m being the mean and δ being the volatility of the logarithm of the percentage jump size. For this lognormal distribution of J, Merton [71] gave a semi-analytical solution to the European vanilla option

$$V = \sum_{n=0}^{\infty} \frac{e^{-\lambda'\tau}\left(\lambda'\tau\right)^n}{n!} V_n^{BS} \tag{5.5.6}$$

where $\lambda' = \lambda(1+k)$, $k = \exp\left(m + \delta^2/2\right) - 1$ is the average jump size measured as a percentage of the asset price, V_n^{BS} is the Black-Scholes price with the risk-free rate r_d adjusted to $r_d - \lambda k + n\ln(1+k)/\tau$ and the Black-Scholes implied volatility adjusted to $\sqrt{\sigma^2 + n\delta^2/\tau}$.

Figure (5.5.17) displays the comparison of Black-Scholes implied vol smiles of the Merton Jump and VV models. Figure (5.5.18) displays the relative error between the Merton Jump and VV prices;put prices are used before K_2 and call prices are used after K_2 to

[1]See [48, 52, 33] for the Merton Jump Diffusion.

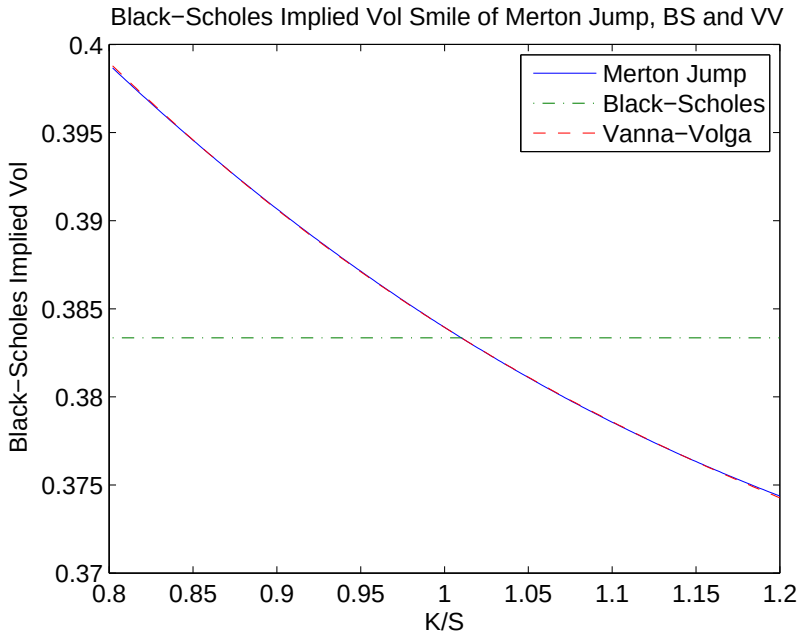

FIGURE 5.5.17. Black-Scholes Implied Vol Smile of Merton Jump, Black-Scholes and VV

compute the relative error. Figure (5.5.19) assumes $m = 0$ and Figure (5.5.20) assumes $\delta = 0.3$. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The parameters for the Merton Jump model and the option are $\sigma = 0.25$, $\tau = 1$, $S = 5$, $r_d = 0.03$, and $r_f = 0.0$. For the ranges of $0 \leqslant \delta \leqslant 1$, $-1 \leqslant m \leqslant 1$, and $0 \leqslant \lambda \leqslant 15$, Figures (5.5.19) and (5.5.20) show that the relative errors between the Merton Jump Model and VV prices are in the order of 10^{-4}. All these test results suggest that the VV option price is a sound alternative to the price obtained through the Merton Jump Model.

5.5.6. Gram-Charlier as the Market. The VV price does not seem to agree as closely with the Gram-Charlier price as with the Heston, SABR, CEV, Merton Jump prices. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The parameters for the Gram-Charlier model and the option are $\tau = 1$, $S = 5$, $r_d = 0.03$, $r_f = 0.0$, $\sigma_{GC} = 0.2$, $\mu_3 = -1.2$ and $\mu_4 = 5$. See the earlier section about the Gram-Charlier model for the details about this model. Figure (5.5.21) shows that the Black-Scholes implied smile given by the VV model is less fitted to the smile given by the Gram-Charlier model than the smiles given by the Heston, SABR, CEV, Merton Jump models. Figure (5.5.22) shows that the relative error between the VV and Gram-Charlier prices is around 2% between K_1 and K_3.

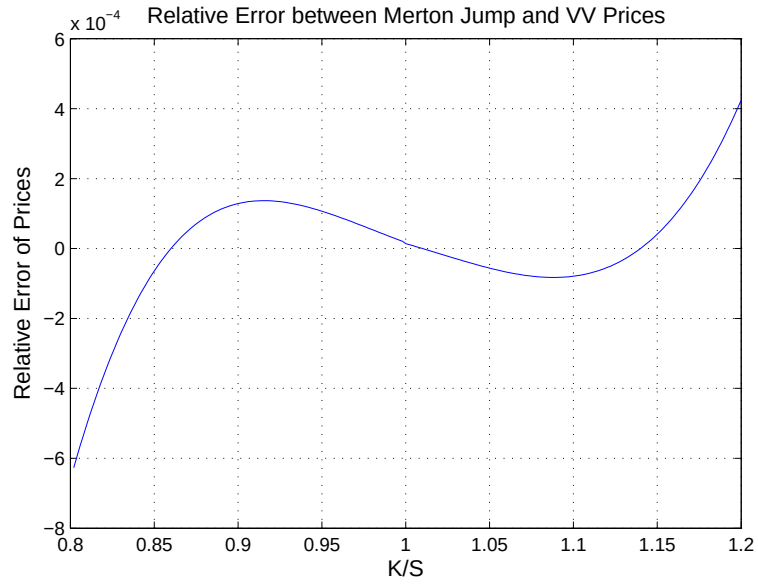

FIGURE 5.5.18. Relative Error between Merton Jump and VV Prices

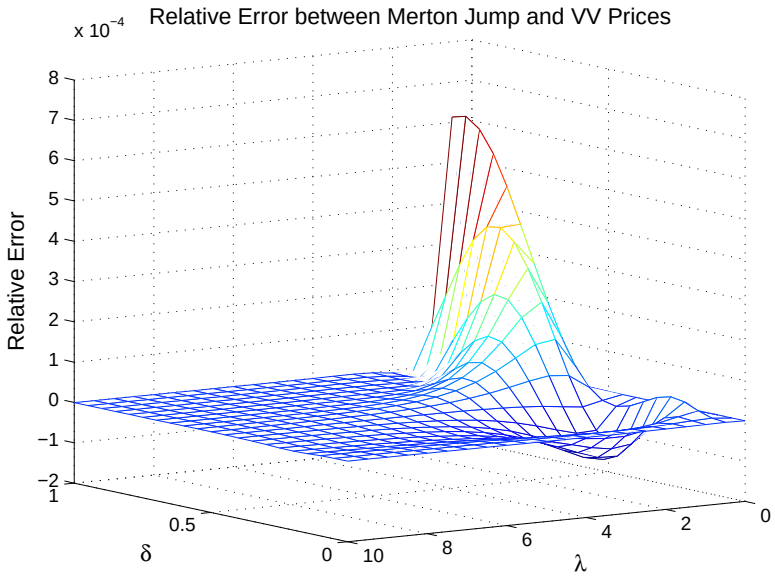

FIGURE 5.5.19. Relative Error between Merton Jump and VV Prices

5.5.7. A Study of the VV Method as a Useful Hedging Tool. We have shown that the VV method is not an arbitrage-free method. Although the VV method is not a perfect hedging tool theoretically, we are interested in knowing whether the VV weights can be used to provide a good hedge in practice.

To serve this purpose, a test is done with the market being assumed to be a double-Heston model (defined in (4.1.1)) and the instrument assumed to be a call option. Assume

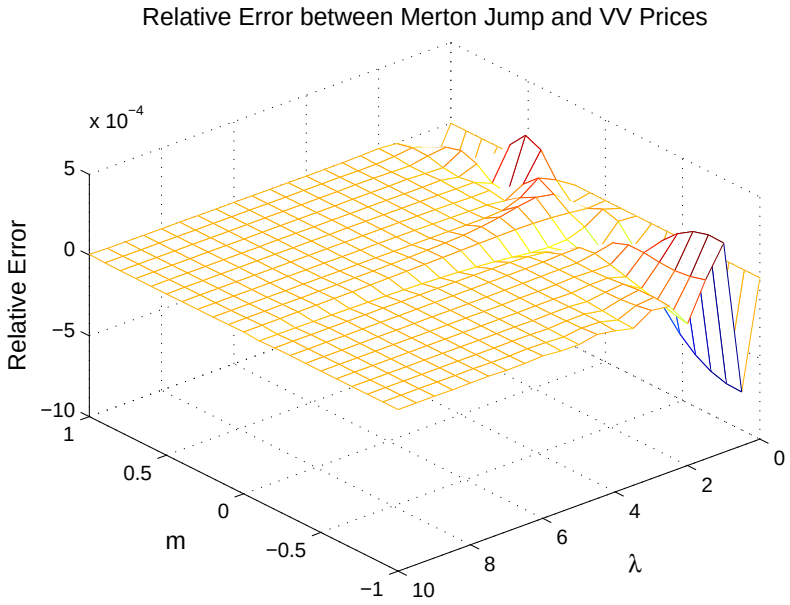

FIGURE 5.5.20. Relative Error between Merton Jump and VV Prices

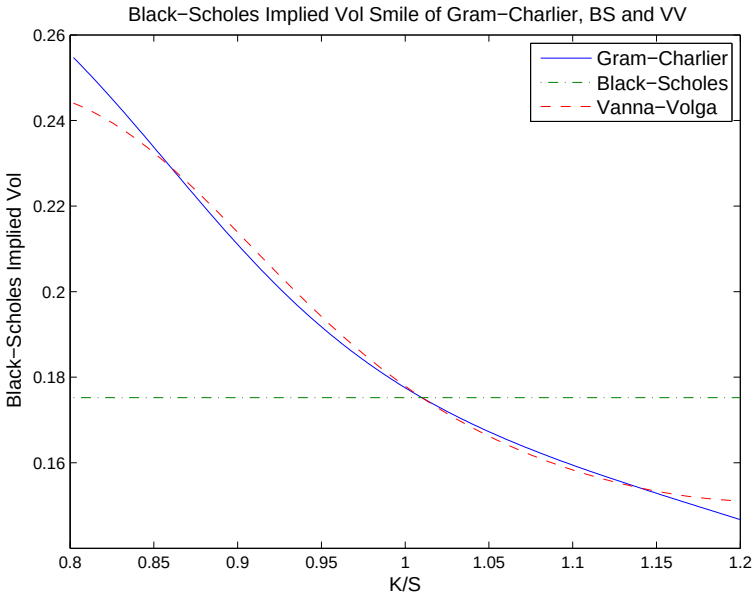

FIGURE 5.5.21. Black-Scholes Implied Vol Smile of Gram-Charlier, BS and VV

$\kappa_1 = 2.13$, $\kappa_2 = 5.49$, $\theta_1 = 0.085$, $\theta_2 = 0.058$, $\xi_1 = 0.27$, $\xi_2 = 0.1$, $\nu_1(0) = 0.091$, $\nu_2(0) = 0.065$, $\rho_1 = -0.86$, $\rho_2 = 0.96$, $r_d = 0$, $r_f = 0$, $S = 5.0$, $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$, $K = 5.35$, and $T = 0.25$ as the testing parameters. The number of the Monte Carlo simulation paths is $100,000$.

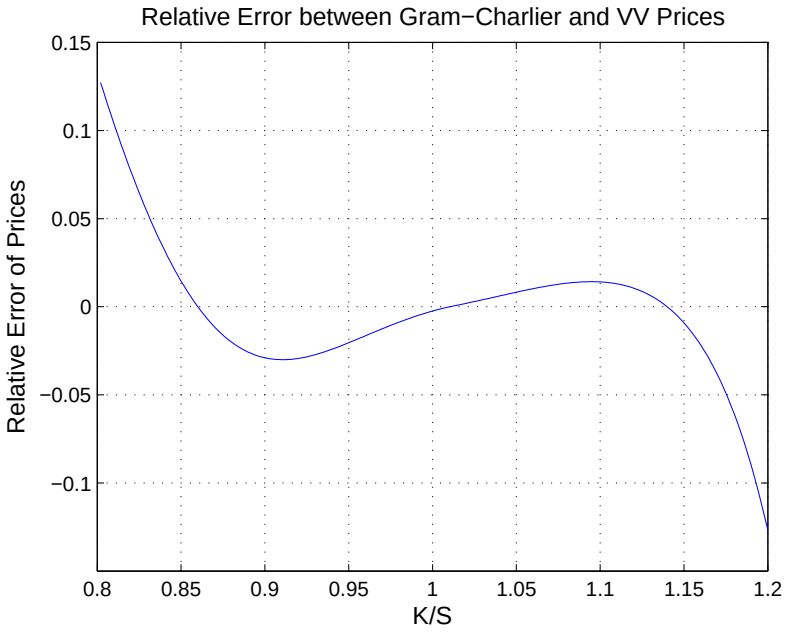

FIGURE 5.5.22. Relative Error between Gram-Charlier and VV Prices

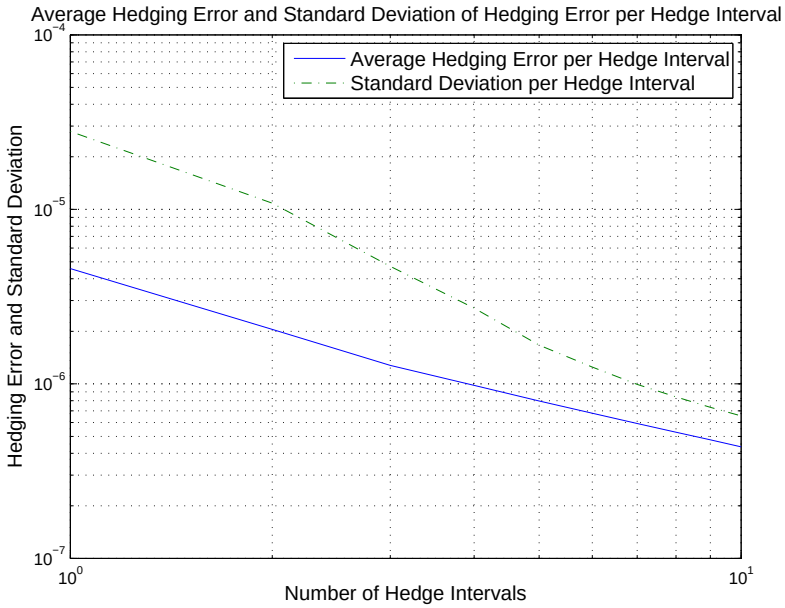

FIGURE 5.5.23. Average Hedging Error and Standard Deviation of Hedging Error per Hedge Interval with the VV Weights in a Double-Heston Market. (See Text for Parameter Values)

The hedging error per hedge interval is computed by $\delta V^{mkt} - \Delta\delta S - \sum_k \omega_k \delta U_k^{mkt}$ (Note the hedging error is $d\Pi - r\Pi dt$, where $\Pi = V^{mkt} - \Delta S - \sum_k \omega_k U_k^{mkt}$; but we set the interest rates zero so that $r\Pi dt$ is zero, so the hedging error $d\Pi$ is expected to be zero in the Black-Scholes model world.), where the market option values are computed with the

double-Heston model and the hedging weights ω_k and Δ are computed by the VV weights (5.2.15) and the delta hedge (5.2.35) in the Black-Scholes world. The actual hedging error in the simulated double-Heston market can reveal how well the VV weights perform as valid hedging strategies. For a three-months dated vanilla call option, Figure (5.5.23) shows that the hedging error per hedge interval is decreasing in the first order as the number of hedge intervals increases (corresponding to an error of $O(\delta t)$, where δt is the size of hedge interval), and the standard deviation of the hedging error per hedge interval is decreasing approximately in the order of $O\left((\delta t)^{3/2}\right)$. Furthermore, for this test, the VV weights demonstrate a superb hedging precision that is very close to static hedging; when there is just a single hedging interval, Figure (5.5.23) gives very good hedging error already; when the number of hedging interval is just 10, the hedging is in very high accuracy.

Because of the demonstrated very low hedging error and the very narrow distribution of the hedging error, at least for the vanillas, the VV weights have the potential to be an excellent hedging tool.

5.5.8. Summary of Key Findings for Vanillas. (1) Based on our numerical analysis, the VV price can arguably be an excellent alternative to the Heston, SABR, CEV, Merton Jump prices for vanilla options because the VV price is very close to the Heston, SABR, CEV, Merton Jump prices. The Heston, SABR, CEV, Merton Jump models must be calibrated. The VV method requires no time-consuming calibration, which is a substantial advantage. The agreement between the VV and Heston, SABR, CEV, Merton Jump prices is true for a wide range of parameters. Although the VV formula is proved roughly the same as the Heston formula only for small ξ and ρ as Polishchuk et al [82] suggested, the VV and Heston prices remain very close for a wide range of ξ and ρ values. (2) The RRBF-VV price is the same as the VV price; the Traders' Rule price is just an approximation to the RRBF-VV price. (3) The VV price is very flexible in modelling different smiles. We know that the Heston, SABR, CEV, Merton Jump model can not be fit to any vol smile; each of these models is only good for certain types of smiles. But the VV method can always be fitted to the three liquid strikes, suggesting the VV method is good for a wide range of market conditions. (4) It is important to use three hedging vanillas U_k in the VV formula. Using only one or two hedging vanillas will result in a substantial increase of error; using only one hedging vanilla will make the VV method fit one market price and using only two hedging vanillas will make the VV price fit only two market prices.

With perfect estimates of market prices on the liquid strikes by the VV method, on other strikes there are underestimates and overestimates. (5) The errors between the VV and Heston, SABR, CEV, Merton Jump prices outside the range between K_1 and K_3 increase quickly as the strike is further away from this range; it suggests that the interpolation of the VV method has better results than its extrapolation. (6) For vanillas, the VV method will always match the market prices on the liquid strikes. For exotics, however, there is no guarantee that for certain conditions the VV method will produce a perfect match to the market price because no market exotic prices can be observed. For example, an exotic option with a fixed strike parameter priced by the VV method on the liquid strikes will not necessarily give the market prices. (7) The VV weights may be excellent hedging strategies.

5.6. The VV Method for Exotics

The VV method is empirically believed to produce excellent mark-to-market vanilla prices. Our analysis above has revealed that the VV method is a highly accurate alternative to the Heston, CEV, SABR, and Merton Jump models. While the calibration process can be time-consuming for the Heston, CEV, SABR, and Merton Jump models, the VV price incurs no calibration cost. However, for exotics, there is no consensus whether and when the VV method has as good performance on exotics among the traders. While the VV price is often the default pricing method for vanillas, the VV price is often one of the pricing choices for exotics.

For different exotics, the VV method is not equally received in banks. One type of exotics for which the VV method is often used in banks is barrier options. Also because the VV method used with barrier options has an additional correctional factor in the VV formula, we will focus on the barrier options in the next chapter.

In this chapter, we will analyse what types of exotics are likely to benefit from the VV method and what types are not, and the reason for that difference.

5.6.1. Unbalanced Vanna Terms - Breakdown of the Hedging Analysis for Path Dependent Exotics. In the previous hedging analysis, we take the Itô's expansion for the portfolio $dV^{BS} - \Delta dS - \sum_k \omega dU_k^{BS}$ with respect to current spot S and the ATM Black-Scholes implied volatility; then in order to cancel out the stochastic terms in the expansion, we arrive at the formula to compute the VV weights ω_k defined in (5.2.15). For vanillas, taking the Itô's expansion this way is fine.

As we point out earlier in the discussion about the implicit assumptions for the hedging analysis, for the path dependent exotic options, this analysis may not apply very well. For instance, when the Black-Scholes exotic option price V is determined not only by the current spot S but also by a function of a history of spot prices (for Asian options this function is the average of spot prices $\overline{S}$; for lookback options this function is the minimum S_{min} or maximum S_{max} price), the VV weights ω_k will not be enough to cancel out all of the stochastic terms in the Itô's expansion because there will be the unbalanced Vanna terms in the Itô's expansion. Take Asian options for example, the unbalanced Vanna term is $\left.\frac{\partial^2 V^{BS}}{\partial \overline{S} \partial \sigma}\right|_{\sigma_{ATM}} d\overline{S} d\sigma_{ATM}$ because the Vannas of the vanillas w.r.t. $\overline{S}$, $\left.\frac{\partial^2 U_k^{BS}}{\partial \overline{S} \partial \sigma}\right|_{\sigma_{ATM}}$ are zero as they are independent of $\overline{S}$. In the Itô's expansion, there are also $(...)\, d\overline{S}$ and $(...)\left(d\overline{S}\right)^2$ terms but they are part of the dt term; that is because from $\overline{S} = \frac{1}{t}\int_0^t S dt$ we obtain $d\left(t\overline{S}\right) = Sdt$, hence $d\overline{S} = \frac{S - \overline{S}}{t} dt$. So the unbalanced Vanna terms are the main issue affecting the accuracy of the VV price for path dependent exotic options. As shown in Figures (5.5.8) and (5.5.9), dropping or replacing the Vanna in the VV formula to compute ω_k will result in large errors. In this example, we discuss the $d\overline{S} d\sigma_{ATM}$ term because the Asian option price (whether it is in the real market or in the Black-Scholes world) depends on the average price $\overline{S}$ from the contract starting time to the present time as well as depending on the current price level S, and to keep the $dSd\sigma_{ATM}$ term balanced will not be enough. The Vanna-Volga method has not been known a popular method for Asian options.

Just like the exclusion of the dS error terms, using the same type of the error analysis as in (5.4.1) will not include the $d\overline{S}$ error terms because the present time $\overline{S}$ is observable. In a flat-vol market, the error convergence order with respect to the Black-Scholes implied volatility will also be like (5.4.2), $\delta V \approx \sum_k \omega_k \delta U_k + O\left((\delta\sigma)^3\right)$. Like the dS error terms, the $d\overline{S}$ error terms make sense in a hedging perspective.

Therefore, from the hedging argument, it can be argued that the VV method will not work as well for the path dependent exotic options as for the options with a single current spot dependence; thus digital options, depending only on the current spot in the Black-Scholes model, should be a very good candidate for the VV method.

Among the path dependent exotic options, it can also be argued that the VV method is more applicable to some exotic options than to the others. The VV method should be more applicable to barrier options than to Asian or lookback options because the knockout

barriers depend on the current spot just like vanillas unless the barriers are hit while the Asian or lookback options always depend on a history of spot.

5.6.2. Performance of the VV Method on Digital Options. The digital options are not path dependent. These options do not have the problem of unbalanced Vanna terms in the Itô's expansion for the portfolio $dV^{BS} - \Delta dS - \sum_k \omega dU_k^{BS}$. According to our analysis, these options should work very well with the VV method. To verify our analysis, we will apply the VV method to the cash-or-nothing options in a market assumed to follow the Heston model.

A cash-or-nothing call has the following payoff :

$$\begin{cases} 1 & S \geqslant K \\ 0 & S < K \end{cases} \text{ for call options, } \begin{cases} 0 & S > K \\ 1 & S \leqslant K \end{cases} \text{ for put options.}$$

The Black-Scholes closed-form formula [48] for cash-or-nothing options is

$$\begin{cases} e^{-r_d T}\mathcal{N}(d_2), & \text{for call options,} \\ e^{-r_d T}\mathcal{N}(-d_2), & \text{for put options.} \end{cases}$$

where $\mathcal{N}$ is the cumulative normal function (3.2.4) and d_2 is defined in (3.2.6). The Heston price for cash-or-nothing calls and puts can also be computed analytically. [65, 89]

The parameter values assumed in the test of cash-or-nothing options for Figures (5.6.1) and (5.6.2) are the same as those assumed in the test of vanillas for Figures (5.5.3), (5.5.4), and (5.5.5). Figure (5.6.1) shows the relative error between the Heston and VV prices for Cash-or-Nothing options at different strikes. When $K > S$ the call prices are compared and otherwise the put prices are compared. Figure (5.6.2) shows that the relative error between the Heston and VV prices for a Cash-or-Nothing call option at $K = 0.5(K_2 + K_3)$ with different Heston model parameters.

The testing conditions are made exactly the same for the digitals as for the vanillas because we want to find whether or how the difference between the VV and Heston prices changes. By a comparison between the test of cash-or-nothing options and the test of vanillas, we can see the VV price is almost as accurate for the cash-or-nothing options as for the vanillas; across different strikes, the relative errors in both tests are in the order of 10^{-3}; under the same model parameters, the relative errors stay in the same order as well.

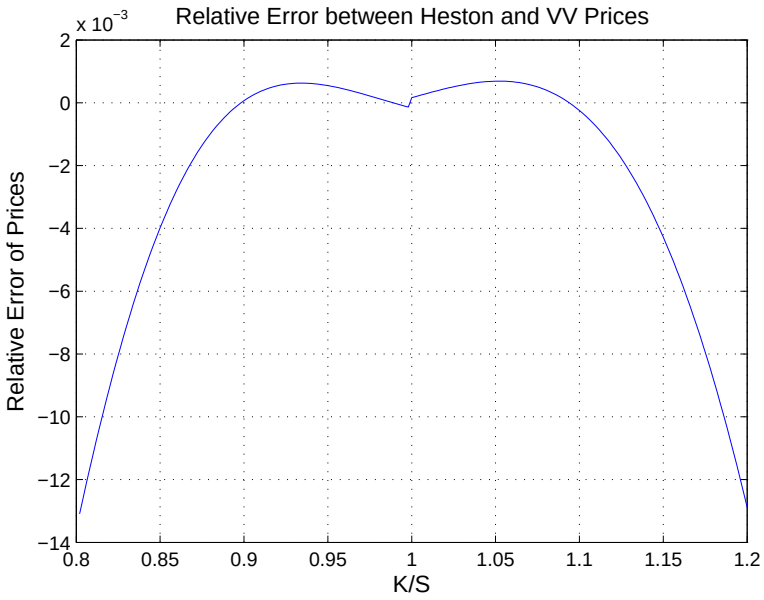

FIGURE 5.6.1. Relative Error between Heston and VV Prices for Cash-or-Nothing Option

The confirms our early analysis that the VV method should work much better when the option is not path dependent.

5.6.3. The Smile-VV Method for Exotics with a Fixed Strike. For an exotic option with a fixed strike parameter, we will explore an alternative way of using the VV method with the following steps:

- Compute the vanilla prices at each strike K using the VV method
- Obtain the Black-Scholes implied vol at each K
- Use the Black-Scholes model to compute the exotic price at strike K using the Black-Scholes implied vol at K

In pricing some path dependent exotics, this method can have an advantage, which is that the Black-Scholes implied vol smile is the same for the vanillas and the exotics. Since the Black-Scholes implied vol smile obtained with the VV method is very good at matching the market smile, it is reasonable to explore the possibility that the Black-Scholes model may be able to produce good prices for some exotics with this smile, provided that the exotic has a fixed strike.

For easy reference, we will call this method the Smile-VV method.

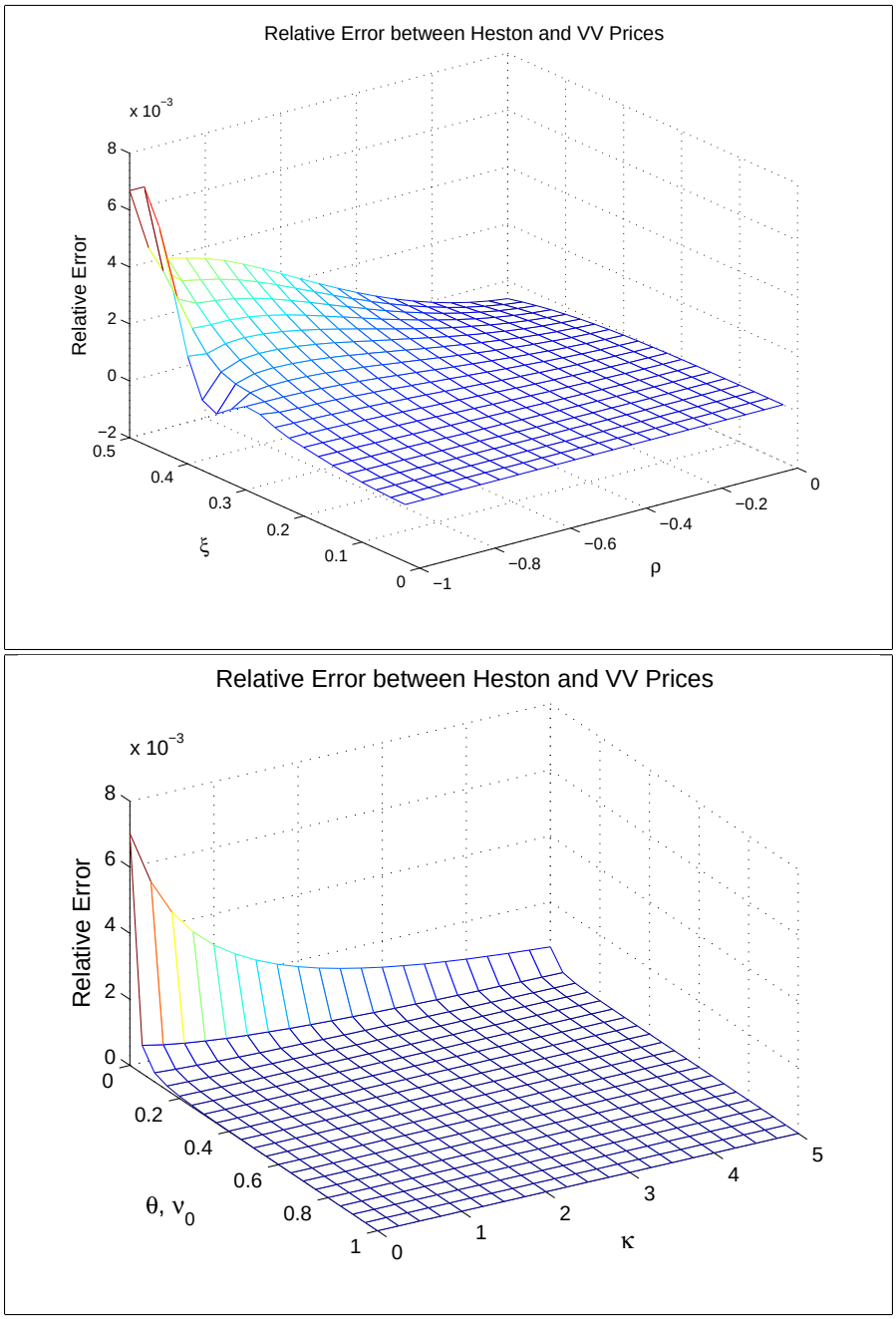

FIGURE 5.6.2. Relative Error between Heston and VV Prices for Cash-or-Nothing Option

It must be emphasised that this method is not intended for all path dependent exotics. In the next section, we will explain why the Smile-VV method may be used as an approximation for Asian options in some cases.

5.6.4. Heston as Market for Fixed Strike Continuous Asian Option. The payoff of a fixed strike continuous Asian option is $\left(\overline{S}-K\right)^{+}$ for a call and $\left(K-\overline{S}\right)^{+}$

for a put where $\overline{S} = \frac{1}{T}\int_0^T S\,dt$ for an arithmetic continuous Asian option and $\overline{S} = \exp\left(\frac{1}{T}\int_0^T \ln S\,dt\right)$ for a geometric continuous Asian option, [61] T is the maturity and K is the strike.

For Asian options, as we will shortly see, the Smile-VV method is shown in examples to perform much better than the Smile-VV method for lookbacks. The Asian option payoff depends on the average underlying asset price from the present time to the maturity. We will use the analysis of the geometric Asian option to provide a simple conceptual explanation for the Smile-VV method.

Assuming the underlying follows the geometric Brownian motion $dS_t = \mu_t S_t dt + \sigma_t S_t dW_t$ where the instantaneous volatility σ_t may be random, Gatheral [36, 63] gave a general path-integral representation of the Black-Scholes implied variance:

$$\sigma_{BS}^2(K, T) = \frac{1}{T}\int_0^T \nu(K, T, t)\,dt \tag{5.6.1}$$

where

$$\nu(K, T, t) \triangleq \mathrm{E}^{G_t}\left[\sigma_t^2\right] \triangleq \frac{\mathrm{E}\left[\sigma_t^2 S_t^2 \Gamma_{BS}(S_t, \overline{\nu_t})\middle|\mathcal{F}_0\right]}{\mathrm{E}\left[S_t^2 \Gamma_{BS}(S_t, \overline{\nu_t})\middle|\mathcal{F}_0\right]}, \tag{5.6.2}$$

in which $\Gamma_{BS}(S_t, \overline{\nu_t})$ is defined as the Black-Scholes gamma of a vanilla option C_{BS} with the variance $\overline{\nu_t}$

$$\Gamma_{BS}(S_t, \overline{\nu_t}) \triangleq \frac{\partial^2}{\partial S_t^2} C_{BS}(S_t, K, \overline{\nu_t}, T-t)$$

and $\overline{\nu_t}$ is defined as

$$\overline{\nu_t} \triangleq \frac{1}{T-t}\int_t^T \nu(K, T, u)\,du$$

Because the definition of $\overline{\nu_t}$ relies on the history of $\nu(K, T, t)$, it is clear that (5.6.2) is an implicit equation about $\nu(K, T, t)$.

The variance of $\ln \overline{S}\big|_{\text{Geometric}}$ satisfies

$$\begin{aligned}
\text{Var}\left[\ln \overline{S}\big|_{\text{Geometric}}\right] &= \text{Var}\left[\frac{1}{T}\int_0^T \ln S_t dt\right] \\
&= \text{Var}\left[\frac{1}{T}\int_0^T \left[\ln S_0 + \int_0^t \left(r - \frac{1}{2}\sigma_s^2\right) ds + \int_0^t \sigma_s dW_s\right] dt\right] \\
&= \text{Var}\left[\ln S_0 + \frac{1}{T}\int_0^T (T-t)\left(r - \frac{1}{2}\sigma_t^2\right) dt + \frac{1}{T}\int_0^T (T-t)\,\sigma_t dW_t\right] \\
&= \text{Var}\left[\frac{1}{T}\int_0^T (T-t)\left(-\frac{1}{2}\sigma_t^2\right) dt + \frac{1}{T}\int_0^T (T-t)\,\sigma_t dW_t\right]
\end{aligned} \tag{5.6.3}$$

For simplicity of discussion, assume that the instantaneous volatility σ_t is only a function of time t, i.e. $\sigma_t = \sigma(t)$. With this assumption, the definition (5.6.2) gives $\sigma_t^2 = \sigma^2(t) = \nu(K, T, t)$ (see Gatheral [36]), and hence from (5.6.3) the variance for $\ln \overline{S}\big|_{\text{Geometric}}$ can be simplified:

$$\text{Var}\left[\ln \overline{S}\big|_{\text{Geometric}}\right] = \frac{1}{T^2}\int_0^T (T-t)^2\,\nu(K, T, t)\,dt. \tag{5.6.4}$$

If $\nu(K, T, t)$ is a constant (which is equal to the Black-Scholes variance $\sigma_{BS}^2(K, T)$ from (5.6.1)) with respect to time, the expression for the variance for $\ln \overline{S}\big|_{\text{Geometric}}$, (5.6.4) becomes $\frac{1}{T^2}\int_0^T (T-t)^2\,\nu(K, T, t)\,dt = \sigma_{BS}^2(K, T)\,T/3$, hence

$$\sigma_{BS}^2(K, T) = \frac{3}{T}\text{Var}\left[\ln \overline{S}\big|_{\text{Geometric}}\right] = \frac{1}{T}\int_0^T \frac{3}{T^2}\nu(K, T, t)\,(T-t)^2\,dt \tag{5.6.5}$$

The right hand side of (5.6.5) can be viewed as the Black-Scholes geometric Asian implied variance. Following the form of (5.6.1), we define

$$\sigma_{BS}^2(K, T)\big|_{\text{Geometric}} = \frac{1}{T}\int_0^T \nu^{\text{Geometric}}(K, T, t)\,dt \tag{5.6.6}$$

where

$$\nu^{\text{Geometric}}(K, T, t) = \frac{3}{T^2}\nu(K, T, t)\,(T-t)^2 \tag{5.6.7}$$

So we now see that the reason why we introduced $\nu(K, T, t)$ from the beginning instead of just discussing $\sigma_{BS}^2(K, T)$ is because we wanted to work out an equivalent version of

$\nu(K, T, t)$ implied by Asian option, $\nu^{\text{Geometric}}(K, T, t)$, and through this we can compare how the Black-Scholes vanilla and Asian implied variances differ.

When $\nu(K, T, t)$ is a constant, we see from above that the Black-Scholes vanilla implied variance is equal to the Black-Scholes geometric Asian implied variance, hence $\sigma_{BS}^2(K, T) = \sigma_{BS}^2(K, T)\big|_{\text{Geometric}}$.

When $\nu(K, T, t)$ is not a constant, the equality (5.6.5) will not necessarily hold. In this case, we will compute the difference

$$\begin{aligned}
&\sigma_{BS}^2(K, T) - \sigma_{BS}^2\big|_{\text{Geometric}}(K, T) \\
&= \frac{1}{T}\int_0^T \nu(K, T, t)\, dt - \frac{1}{T}\int_0^T \frac{3}{T^2}\nu(K, T, t)(T-t)^2\, dt \\
&= \frac{1}{T^3}\int_0^T \nu(K, T, t)\left[T^2 - 3(T-t)^2\right] dt
\end{aligned} \tag{5.6.8}$$

Assuming $\nu(K, T, t)$ takes the simple term structure form based on the Taylor expansion up to the quadratic terms of $\nu(K, T, t) = ae^{-\lambda t} + \theta$ (which through (5.6.1) gives an exponential $\sigma_{BS}^2(K, T)$ with respect to T (see [36] for this assumption)), i.e. $\nu(K, T, t) \approx a\left(1 - \lambda t + \frac{1}{2}\lambda^2 t^2\right) + \theta$, we will compute the error of $\sigma_{BS}^2(K, T) - \sigma_{BS}^2\big|_{\text{Geometric}}(K, T)$. This term structure and (5.6.8) give

$$\begin{aligned}
&\sigma_{BS}^2(K, T) - \sigma_{BS}^2\big|_{\text{Geometric}}(K, T) \\
&= \frac{1}{T^3}\int_0^T \left[a\left(1 - \lambda t + \frac{1}{2}\lambda^2 t^2\right) + \theta\right]\left[T^2 - 3(T-t)^2\right] dt \\
&= -\frac{1}{4}a\lambda T^4 + \frac{43}{60}a\lambda^2 T^5
\end{aligned} \tag{5.6.9}$$

For the case in which σ_t is only a function of time t, therefore, when the term structure decaying component is small (a is small) or there is not much decay (λ is small), or the maturity is short-dated ($T \ll 1$), the Black-Scholes vanilla implied variance σ_{BS}^2 can be used to approximate the Black-Scholes Asian implied variance $\sigma_{BS}^2\big|_{\text{Geometric}}$.

For the general case where σ_t is not just a function of time t, and rather it can be random or a function of S_t, the computation of $\text{Var}\left[\ln \overline{S}\big|_{\text{Geometric}}\right]$ using (5.6.3) is rather complex; as a consequence, the computation of $\nu^{\text{Geometric}}(K, T, t)$ is difficult too. In this case, in addition to a term structure, $\nu(K, T, t)$ and $\nu^{\text{Geometric}}(K, T, t)$ should have a smile with respect to the strike K. We shall for this case provide a crude qualitative analysis to get some intuition.

Assume that $\nu(K, T, t) = a(K, T) e^{-\lambda(K,T)t} + \theta(K, T)$. Assume that $\nu^{\text{Geometric}}(K, T, t)$ also follows an exponential term structure but has different decaying rate and smile from $\nu(K, T, t)$, i.e. $\nu^{\text{Geometric}}(K, T, t) = a^*(K, T) e^{-\lambda^*(K,T)t} + \theta^*(K, T)$. Therefore, we obtain

$$\begin{aligned}
&\sigma_{BS}^2(K, T) - \sigma_{BS}^2\big|_{\text{Geometric}}(K, T) \\
&= \frac{1}{T}\int_0^T a(K, T) e^{-\lambda(K,T)t} + \theta(K, T)\, dt \\
&\quad -\frac{1}{T}\int_0^T a^*(K, T) e^{-\lambda^*(K,T)t} + \theta^*(K, T)\, dt \\
&= \frac{1}{T}\left[a(K, T)\frac{1 - e^{-\lambda(K,T)T}}{\lambda(K, T)} - a^*(K, T)\frac{1 - e^{-\lambda^*(K,T)T}}{\lambda^*(K, T)}\right] \\
&\quad +\theta(K, T) - \theta^*(K, T) \\
&\approx -\frac{1}{2}(a(K, T) - a^*(K, T))\lambda(K, T) T \\
&\quad a(K, T) + \theta(K, T) - a^*(K, T) - \theta^*(K, T) \\
&\quad +O\left(a(K, T)\lambda^2(K, T) T^2, a^*(K, T)(\lambda^*(K, T))^2 T^2\right)
\end{aligned} \tag{5.6.10}$$

When the term structure decaying component is small ($a(K, T)$ and $a^*(K, T)$ are small) or there is not much decay ($\lambda(K, T)$ and $\lambda^*(K, T)$ are small), or the maturity is short-dated ($T \ll 1$), $\sigma_{BS}^2(K, T) - \sigma_{BS}^2\big|_{\text{Geometric}}(K, T)$ becomes smaller with the elimination of the first term in (5.6.10) ($-\frac{1}{2}(a(K, T) - a^*(K, T))\lambda(K, T) T$), and approaches $a(K, T) + \theta(K, T) - a^*(K, T) - \theta^*(K, T)$. With the further assumption that $a(K, T) + \theta(K, T) - a^*(K, T) - \theta^*(K, T)$ is small, the Black-Scholes vanilla implied variance σ_{BS}^2 can be used to approximate the Black-Scholes Asian implied variance $\sigma_{BS}^2\big|_{\text{Geometric}}$.

For lookbacks, the above analysis does not apply; thus we will later see that the Smile-VV method performs not as well with lookbacks as Asians.

The Black-Scholes model price for the fixed strike arithmetic continuous Asian option is computed through the PDE method developed by Vecer (2001[97], 2002[98]). The Vecer approach assumes the same log-normal model (the spot follows the process (5.1.2).) for the underlying as the Black-Scholes model. Through the Vecer method, a fixed strike Asian call price option can be obtained by $V = S_0 U$, where S_0 is the current spot level, and U is the solution of the following PDE and terminal condition,

$$\begin{aligned} U_t + \frac{1}{2}\left(z - e^{-r_f t} q_t\right)^2 \sigma^2 U_{zz} &= 0 \\ U(T, z) &= (z)^+ \end{aligned} \tag{5.6.11}$$

where

$$\begin{aligned} q_t &= \frac{1}{(r_d - r_f)\,T}\left(e^{-r_f(T-t)} - e^{-r_d(T-t)}\right) \\ z_0 &= \frac{1}{(r_d - r_f)\,T}\left(e^{-r_f T} - e^{-r_d T}\right) - e^{-r_d T}\frac{K}{S_0} \end{aligned} \tag{5.6.12}$$

The Black-Scholes model price for the fixed strike geometric continuous Asian option is computed by the analytical solution given by Kemna and Vorst: [57]

$$\begin{aligned} V_{call} &= Se^{(b-r_d)(T-t)-}\mathcal{N}(d_1) - Ke^{-r_d(T-t)}\mathcal{N}(d_2) \\ V_{put} &= -Se^{(b-r_d)(T-t)-}\mathcal{N}(-d_1) + Ke^{-r_d(T-t)}\mathcal{N}(-d_2) \end{aligned}$$

where

$$d_1 = \frac{\log(S/K) + (b + \frac{1}{2}\sigma_A^2)(T-t)}{\sigma_A(\sqrt{T-t})},\ d_2 = \frac{\log(S/K) + (b - \frac{1}{2}\sigma_A^2)(T-t)}{\sigma_A(\sqrt{T-t})}$$

and

$$\sigma_A = \sigma/\sqrt{3},\ b = \frac{1}{2}\left(r_d - r_f - \frac{\sigma^2}{6}\right)$$

The market Asian option price and the market vanilla prices are assumed to follow the Heston model. The Heston Asian prices are computed by Euler Monte Carlo simulations. The market vanilla prices are computed with the closed-form Heston formula. The parameters for the Heston model are $\kappa = 1.1$, $\theta = 0.09$ $\xi = 0.27$, $\rho = -0.7$, $\nu_0 = 0.09$, $\tau = 0.6$, $S = 5$, $r_d = 0.03$ and $r_f = 0.0$. The maturity for the Asian and vanillas is $T = 0.6$. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The PDE (5.6.11) is solved with the code by Lehalle.[64] Vecer in his papers presented results comparing with several different methods such as Monte Carlo for the no dividends case, which the code well replicates. Over $z \in [-1, 1]$ and $t \in [0, T]$, both the numbers of spatial grid points and time steps are set to 500. For the Heston simulations, the number of time steps is set to 500 and the simulation path number is set to one million. Under these parameters, Figure (5.6.3) shows the errors and relative errors between the Heston call prices and the VV call prices for different strikes; the errors and relative errors between the Heston call prices and the Black-Scholes call prices are also included for comparison. The Black-Scholes Asian call option prices are computed by using the Black-Scholes implied vol at K_2. Figure (5.6.3) shows that the VV prices result in larger errors than the

Black-Scholes prices for most strikes but the Smile-VV prices have much smaller errors than the Black-Scholes prices across all strikes.

The Smile-VV method seems to work much better with the Asian options than the VV method. The Smile-VV circumvents the issue of unbalanced Vanna that would be encountered if the VV method is applied directly. Whether in the in-the-money or in the out-of-the-money range of K, the Smile-VV prices achieve much better accuracy than the VV or Black-Scholes prices. The relative errors of the Smile-VV prices are around 1% to 3% for most strikes; the relative errors of the Black-Scholes price are are around six times those of the Smile-VV prices; the relative errors of the VV price are comparable to those of the Black-Scholes prices.

Figure (5.6.4) displays the relative errors between the Heston and VV prices for an arithmetic Asian call at $K = 0.5\,(K_2 + K_3)$ with respect to different Heston parameters ξ and ρ. It shows that the VV prices are close to the Heston prices if either ξ or ρ is small. Especially, when ξ or ρ is close to zero, the relative errors are very small.

Figure (5.6.5) displays the relative errors between the Heston and Smile-VV prices for an arithmetic Asian call at $K = 0.5\,(K_2 + K_3)$ with respect to different Heston parameters ξ and ρ. Compared the VV price errors in Figure (5.6.4), the Smile-VV price errors are much smaller. Under the same parameter values, as shown in (5.6.5), the Smile-VV prices are roughly 10 times more accurate than the VV prices for Asian options. On the other hand, the VV prices show a rapidly increasing error as the strike moves further out of money away from the spot. This suggests that the dynamical behaviour of the VV model appears to disagree with the market dynamics.

We will explain why the relative errors of the VV prices for arithmetic Asian options are very small when ξ and ρ are close to zero as shown in Figure (5.6.4). For Asian options, as we have discussed earlier, the unbalanced Vanna term is $d\overline{S}d\sigma_{ATM}$ in the portfolio $dV^{BS} - \Delta dS - \sum_k \omega dU_k^{BS}$. Because the market follows the Heston model, σ_{ATM} can be viewed as a function of current S, ν, hence $d\sigma_{ATM}$ containing the dS and $d\nu$ terms. Assume $\overline{S} = \frac{1}{n}\sum_{i=0}^{n} S_i$, where S_i is equally sampled $n+1$ times across the time period$[0,\, t]$, S_n being S and t being current time. The continuous Asian option can be viewed as the sampling of S_i becomes extremely frequent. In $d\overline{S}d\sigma_{ATM}$, the $dSd\overline{S} = (dS)^2$ part is just part of the dt term. The $d\overline{S}d\nu$ part in the $d\overline{S}d\sigma_{ATM}$ terms is what is left unbalanced.

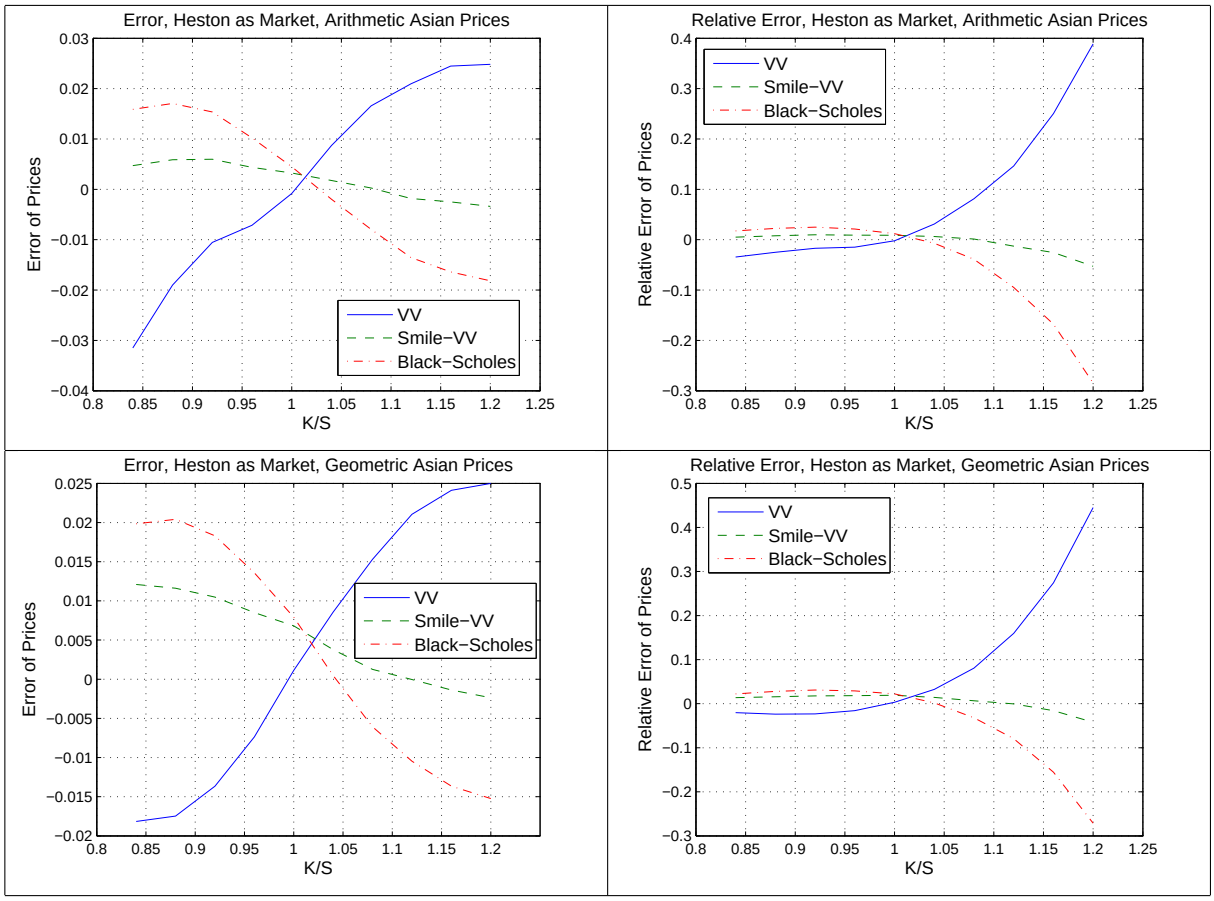

FIGURE 5.6.3. Error (Left) and Relative Error (Right) with the Heston Market for Arithmetic (Top) and Geometric (Bottom) Asian call Option

Now $d\overline{S}d\nu = \frac{1}{n}\sum_{i=1}^{n} dS_i d\nu = \frac{1}{n}\xi\rho S\nu dt$. Note $dS_i d\nu = 0$ if $i < n$. When ξ and ρ are close to zero, the $d\overline{S}d\nu$ term is insignificant in value; therefore, although the term $d\overline{S}d\sigma_{ATM}$ is unbalanced, it will only incur small errors in the VV prices if ξ and ρ are small.

5.6.5. Heston as Market for Lookback Options. The payoff of a fixed strike lookback option is $(S_{max} - K)^+$ for a call and $(K - S_{min})^+$ for a put where $S_{max} = \max_{t\in[0,T]}(S_t)$ and $S_{min} = \min_{t\in[0,T]}(S_t)$, T is the maturity and K is the strike. The payoff of a floating strike lookback option is $(S - S_{min})^+$ for a call and $(S_{max} - S)^+$ for a put. The Black-Scholes model price for the fixed strike lookback option can be computed through the closed form formula developed by Conze & Viswanathan [22]. The Black-Scholes model price for the floating strike lookback option can be computed through the closed form formula developed by Goldman, Sosin & Satto[39].

The parameters for the Heston model are $\kappa = 1.1$, $\theta = 0.09$ $\xi = 0.27$, $\rho = -0.7$, $\nu_0 = 0.09$, $\tau = 0.6$, $S = 5$, $r_d = 0.03$ and $r_f = 0.02$. The maturity for the lookback and

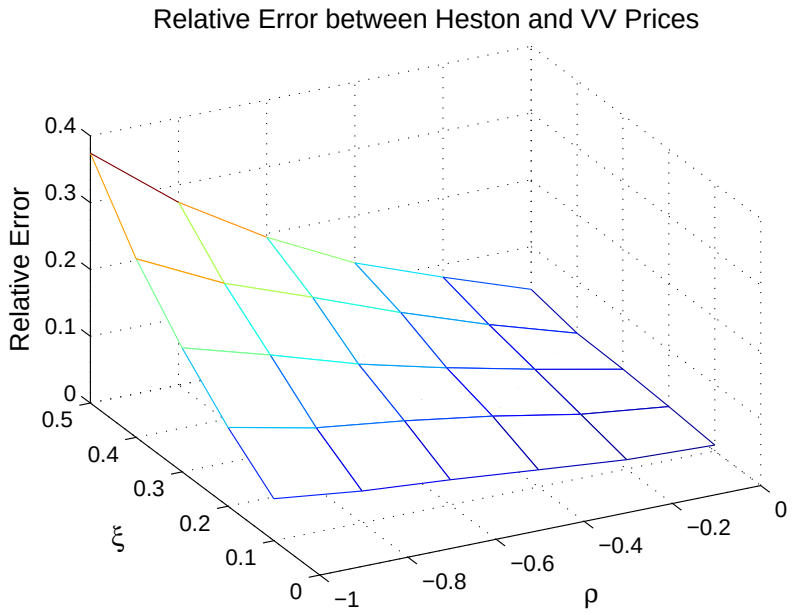

FIGURE 5.6.4. Relative Error between Heston and VV Arithmetic Asian Call Prices w.r.t. ξ and ρ Parameters

FIGURE 5.6.5. Relative Error between Heston and Smile-VV Arithmetic Asian Call Prices w.r.t. ξ and ρ Parameters

vanillas is $T = 0.6$. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. Assume also $t = 0$ at the contract origination time. For the Heston simulations, the number of time steps is set to 500 and the simulation path number is set to one million. The left plot of Figure (5.6.6) shows the comparison of the assumed Heston market, the

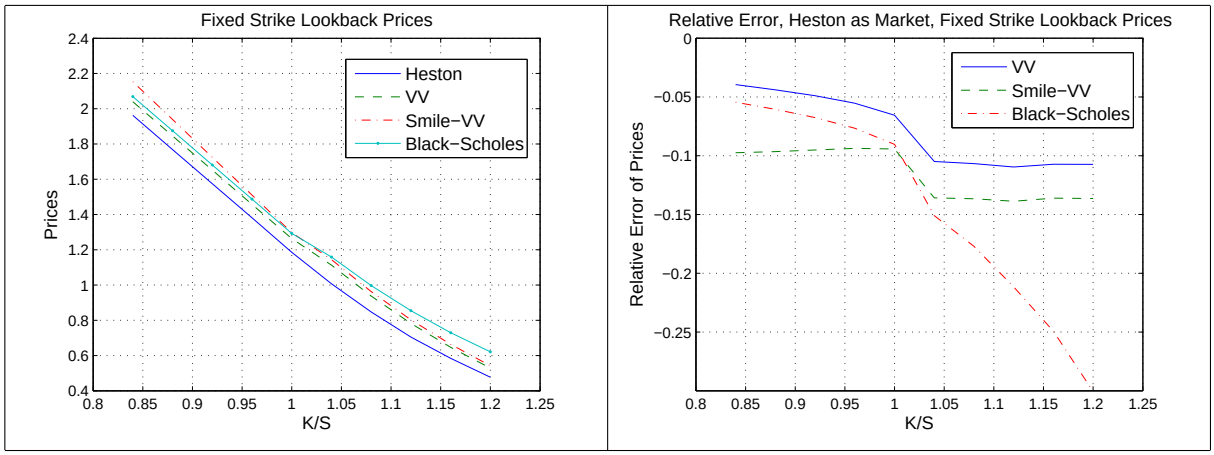

FIGURE 5.6.6. Prices (Left) and Relative Error (Right) with the Heston Market for Fixed Strike Lookback call Option

VV and the Smile-VV and the Black-Scholes call prices at various strikes; the right plot shows the relative errors at various strikes between the assumed Heston market call prices and the call prices of the VV, Smile-VV and Black-Scholes models.

With the parameters set the same as those in the above test of Asian options, Figure (5.6.6) shows that the VV method performs better on lookback options than on Asian options. The relative errors of the VV method for this lookback test stay around 5% for the in-the-money strikes and 10% for the out-of-the-money strikes. The relative errors of the Smile-VV method are around 10% for the in-the-money strikes and 15% for the out-of-the-money strikes. Both the VV and the Smile-VV methods have significant error improvement over the Black-Scholes method. However, compared to path independent options, the VV method does not work as well with the lookback options.

The left plot of Figure (5.6.7) shows the relative errors between the Heston market prices and the VV prices for different ξ and ρ values of the Heston market; the left plot is for a fixed strike lookback call option at $K = 0.5\,(K_2 + K_3)$ and the right plot is for a floating strike lookback call option. Overall, the general trend is that when ξ and ρ are small, the VV prices are close to the Heston prices. Similar to the analysis presented in the last section, for lookback options, the $dS_{max}d\sigma_{ATM}$ or $dS_{min}d\sigma_{ATM}$ term is unbalanced; the unbalanced part in the $dS_{max}d\sigma_{ATM}$ or $dS_{min}d\sigma_{ATM}$ term will contain the $dS_{max}d\nu$ or $dS_{min}d\nu$ term. When ξ and ρ are small, $dS_{max}d\nu$ and $dS_{min}d\nu$ terms are small too.

5.6.6. Summary of Key Findings for Exotics. (1) The VV prices of the path dependent exotic options are much less close to the Heston prices than the VV prices of

FIGURE 5.6.7. Relative Error with the Heston Market for Fixed (Left) and Floating (Right) Strike Lookback call Option

the path independent exotic options such as digitals. The increased errors are related to the unbalanced Vanna terms such as $d\overline{S}d\sigma_{ATM}$, $dS_{max}d\sigma_{ATM}$ and $dS_{min}d\sigma_{ATM}$ in the way of computing the hedging weights ω_k for the VV method. (2) When the Heston parameters ξ and ρ are smaller, these unbalanced Vanna terms become less important in value, making the VV prices closer to the Heston prices. However, even with small ξ and ρ values, the relative errors between the VV and Heston prices for Asians and lookbacks are around 1% to 5% while the relative errors for the vanillas are in the order around 10^{-4} to 10^{-3} for a large range of ξ and ρ values. (3) The Smile-VV method performs much better than the VV method for Asian options; for most ξ and ρ values shown in Figure (5.6.5), the relative errors between the Smile-VV and Heston prices are less than 5%. (4) The VV formula applied to an exotic option does not guarantee to produce market prices for any strikes (if this option does have a strike). (5) Barrier options, which do not have an unbalanced Vanna term problem, should be a good candidate for the VV method. We will discuss barrier options in the next chapter.

CHAPTER 6

The Vanna-Volga Method for Barrier Options

6.1. The P-VV Method

In trading practice, the VV formula for a barrier option is usually adjusted by a factor, denoted by $P \in [0,1]$, so that (5.2.14) becomes

$$V^{P-VV} = V^{BS}\big|_{\sigma_{ATM}} + P\left[\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}\right)\right] \tag{6.1.1}$$

and (5.3.1) becomes

$$\begin{aligned} V^{P-VV} = V^{BS}\big|_{\sigma_{ATM}} &+ P\left[\omega_{RR}\left(\widetilde{RR}^{mkt} - \widetilde{RR}^{BS}\Big|_{\sigma_{ATM}}\right)\right. \\ &+ \left.\omega_{BF}\left(\widetilde{BF}^{mkt} - \widetilde{BF}^{BS}\Big|_{\sigma_{ATM}}\right)\right] \end{aligned} \tag{6.1.2}$$

where

$$P = P_{NoTouch} \text{ is the no touch probability for knock-out Options;} \tag{6.1.3}$$

here ω_k satisfy (5.2.15) and ω_{RR}, ω_{BF} satisfy (5.3.2). The Black-Scholes model prices for the barrier options can be computed analytically: Merton [70] developed a closed form formula for down and out call options; Reiner & Rubinstein [86] provided the closed form formulas for all eight types of barrier options; Haug [41] provided a well organised generalisation of the Reiner & Rubinstein formulas. We will focus on knock-out barrier options in this thesis because a knock-in barrier can be priced with a vanilla and a knock-out option by in-out parity.

The choice of P is not unique in practice. Some practitioners adopt the following P:

$$P = P_{ex} = T_{ex}/T \tag{6.1.4}$$

where T_{ex} is the expected early exercise time defined by

$$T_{ex} \;=\; P_{NoTouch}T + \textstyle\int_0^T t\, p\,(t)\, dt,$$

where $p(t)$ is the probability density for the barrier being hit at t. Hence $P_{ex} = T_{ex}/T > P_{NoTouch}$ always holds. The difference between P_{ex} and $P_{NoTouch}$ is $\frac{1}{T}\int_0^T t\,p(t)\,dt$. Between $P_{NoTouch}$ and P_{ex}, the choice of $P_{NoTouch}$ is more common. Both $P_{NoTouch}$ and P_{ex} are often estimated at the ATM vol using the Black-Scholes model; Note that as a result of using the Black-Scholes model, both P factors are estimated with the expected spot movement under the risk-neutral probability. Bossens et al [9] studied how to used parametrised scaling factors that is based on the commonly used choices $P_{NoTouch}$ or P_{ex}, and systematically calibrate the parameters in the scaling factors. As (6.1.1) is common for knockout barrier options, we will focus our discussion on the form of (6.1.1) in this thesis.

The very existence of, and the specific choices of, P have only some empirical basis but without solid justification. Nevertheless, the FX trading community has a plausible and intuitive explanation as to why some factor P is needed in the VV method. The prevailing intuition is that to keep a hedging portfolio is expensive, and after the exotic option V is already exercised, the hedging portfolio is no longer needed, hence a necessary scale-down of the "Vanna-Volga costs". This intuition only says that a P is possibly needed but does not explain why the amount of the Vanna-Volga costs reduction will be tied to the specific choices of P used in practice.

For easy reference to this form of VV method, we will call this method the P-VV method. For the choice of $P = P_{NoTouch}$, the P-VV method will also be called the $P_{NoTouch}$-VV method; for the choice of $P = P_{ex}$, the P-VV method will also be called the P_{ex}-VV method; the unmodified VV method will be called the VV method.

6.2. Why the P-VV Method is Used

6.2.1. Inconsistency of the VV Price When the Barrier is hit. We will now study the direct source of the requirement for a P factor. When the spot just hits the barrier, the Vega and Volga of V^{BS} become zero but the Vanna does not, making the ω_k non-zero according to the formula for ω_k. The vanilla difference $U_k^{mkt} - U_k^{BS}$ has nothing to do with the barrier and will not be zero except at the ATM vol (which only applies to $U_2^{mkt} - U_2^{BS}$). So when the barrier is hit, the VV price of a knockout barrier $V^{VV} \neq V^{BS} = 0$ holds.

For a knock-out barrier, the market price $V^{mkt} = V^{BS} = 0$ holds when the spot hits the barrier. Hence ω_k would need to be zero at the time the barrier being hit if the VV price were consistent with the market. Therefore, it is a problem that the ω_k in the VV price formula are actually not zero when the barrier is hit. That will lead to wrong prices when the spot is close to the barrier.

6.2.2. Correctness in Two Extreme Ends of P. The use of no touch probability $P_{NoTouch}$ or P_{ex} as the choice of P for knock-out options makes sense in the two extreme cases, either when the underlying is far away from the barrier with virtually no possibility of an early expiry or when the option expires at the current underlying price due to a hit on the barrier. In the former case, $P_{NoTouch} = P_{ex} = 1$; in the latter case, $P_{NoTouch} = P_{ex} = 0$.

When the no touch probability is close to 1, it means this barrier option will expire almost on the maturity T with little possibility of early expiry; this option can be approximately viewed as a vanilla option. The analysis in the last chapter shows that the VV method works very well with vanillas. So the VV price for these barriers can be expressed as

$$V^{P-VV} = V^{BS}\big|_{\sigma_{ATM}} + \left[\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}\right)\right] \tag{6.2.1}$$

When $P = 0$, which means the current underlying price already makes the option expire, the VV price becomes

$$V^{P-VV} = V^{BS}\big|_{\sigma_{ATM}} = 0 \tag{6.2.2}$$

which is a correct price.

The P corrected VV method demonstrates good behaviour in the extreme ends when P is at either 0 or 1. However, it has not been unclear how $P_{NoTouch}$ or P_{ex} performs in the formula when P falls in the middle part of [0, 1].

6.2.3. First Issue of Price Inconsistency at t_h - Liquidating the Hedge When Barrier is Hit. The P-VV method will be discussed using the dynamic hedging argument just like the VV method. Carr and Chou [13] described a static hedging method for barrier options; they suggested that a barrier option can be replicated by vanillas if the vanillas on all strikes are available. Since the FX market has a very limited number of liquid strikes, thus a dynamic hedging strategy is more applicable.

Suppose the barrier is hit at time t_h, $0 < t_h \leqslant T$, where the current time is 0. The VV price of a down-out barrier option will be consistent with the market price if the VV weights ω_k are set to zero at t_h. The setting of ω_k to zero at t_h corresponds to the hedging practice in reality that before t_h, traders use currencies and vanilla options to hedge the risk of the target option; at t_h, traders will liquidate the hedge because the hedge is no longer needed.

At $t = 0$, we do not know if the barrier will be hit before T so we want to hedge the risk of it. At a time t, $0 \leqslant t \leqslant t_h$, the Black-Scholes value growth of the hedged portfolio for the knockout barrier V will follow (5.2.38):

$$\begin{aligned} & dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS} \\ & = r_d \left[V^{BS} - S\Delta - \sum_k \omega_k U_k^{BS} \right] dt + r_f S \Delta dt \end{aligned} \tag{6.2.3}$$

where the Black-Scholes option prices are valued at σ_{ATM}.

At $t = t_h < T$, the task of hedging is not yet finished because we still hold a portfolio of $-\Delta S - \sum_k \omega_k U_k^{BS}$ at this time despite $V^{BS} = 0$. Because the Vanna of V^{BS} is not zero at the time when the barrier is hit, ω_k and Δ are not zero at t_h. The vanillas U_k expire at the maturity T; the portfolio $-\Delta S - \sum_k \omega_k U_k^{BS}$ remains stochastic at t_h; so continuing to use the VV weights (5.2.15) at $t = t_h$ will not make this portfolio hedged. However, (6.2.3) will hold if ω_k and Δ are set to zero at $t = t_h$. Therefore, the modified hedging weights

$$\omega_k = \begin{cases} \text{the VV weights satisfying (5.2.15)}, & 0 \leqslant t < t_h, \\ 0, & t_h \leqslant t \leqslant T. \end{cases} \tag{6.2.4}$$

will keep (6.2.3) well hedged between time 0 and T.

Although the VV price is correct at $t = t_h$ with the modified VV weights (6.2.4), we will describe the second issue that can not be solved by (6.2.4).

6.2.4. Second Issue of Price Inconsistency before t_h - A Probabilistic View to Reduce the Hedging Requirement. The weights computed by (6.2.4) are flawed. Suppose we are pricing a down-out barrier option at time t_h^-, which denotes a time smaller than but infinitesimally close to t_h. Because t_h^- is not yet t_h, the VV price will be computed by (5.2.15); because the Vanna of a down-out barrier option can be significant when the barrier is being approached, there may be a sizable VV correction at t_h^-. So the problem is

that the VV price computed by (6.2.4) may have a not so insignificant value at t_h^-, but will jump to zero at t_h. This implied that the VV price computed at t_h^- based on (6.2.4) may be incorrect. We will analyse this issue and argue that the reason is that by using the VV weights, more hedging is committed than necessary. Although the hedging (6.2.4) may be used by some traders in a real market, it may be too much to correct the Black-Scholes price in the VV method. To keep a hedge carries costs, hence an unnecessary hedge level should be avoided.

The purpose of dynamic hedging is to cancel out the uncertainty in the cashflow of the target option. Maintaining (6.2.3) for each small time step dt is is not the purpose itself but just a means to achieve that purpose. When the VV method is applied to a vanilla (V^{BS} is now expressed by U^{BS}), the vanilla Black-Scholes price U^{BS} needs to be hedged no matter how U^{BS} will evolve because the payoff of the target vanilla is determined by S_T, which is uncertain unless at the maturity. Therefore, maintaining (6.2.3) for each small time step dt until maturity is correct for vanillas. For a down-out barrier option, however, the payoff is always zero if the barrier is hit; with the probability $1 - P_{NoTouch}$, the payoff of a down-out barrier option is certain and predictable; this gives rise to the possibility of reduced hedging. Therefore, just setting ω_k to zero at t_h as described by (6.2.4) is not a good enough remedy to the unmodified VV method.

Let we take another look. We will assume that we have some knowledge about the future movement of the spot. Suppose that we know that the spot movement over $[0, T]$ will take a path that does not hit the barrier, $t_h > T$. The above problem will not exist, so (6.2.3) applies to the entire period of $[0, T]$ and ω_k will be the VV weights defined in (5.2.15).

However, if we expect that the barrier will be hit at $t_h \leqslant T$, we know the current price $V^{BS}\big|_{t=0}$ and the future prices $V^{BS}\big|_{t=t_h} = V^{BS}\big|_{t=T} = 0$; because we know what the price of the barrier option will be at T for sure, we do not need to hedge the price movement of this option at all. In the case of knowing $t_h \leqslant T$, ω_k and Δ should be zero. So in this case, no hedging costs need to be added to the Black-Scholes price.

Of course, we do not know how the evolution of V^{BS} will eventually be in future at time 0. But we have some knowledge about the probability that the barrier will be hit before T. We know that the probability that we will need to add in the VV hedging costs to the ATM Black-Scholes price is $P_{NoTouch}$. So the VV price for barrier options will be

adjusted to the P-VV price:

$$\begin{aligned} V^{P-VV} &= P_{NoTouch}\left(V^{BS}\big|_{\sigma_{ATM}} + \left[\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}\right)\right]\right) \\ &\quad +(1-P_{NoTouch})\left(V^{BS}\big|_{\sigma_{ATM}} + \left[\sum_k 0\left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}\right)\right]\right) \\ &= V^{BS}\big|_{\sigma_{ATM}} + P_{NoTouch}\left[\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}\right)\right] \end{aligned} \tag{6.2.5}$$

This explains the choice of $P = P_{NoTouch}$ for the P-VV method.

As we see from (6.2.5), the estimation error in $P_{NoTouch}$ will contribute to the error of V^{P-VV}. However, we want to note that even if the $P_{NoTouch}$ is perfectly estimated, V^{P-VV} may still be different from the V^{mkt}. As we have analysed, the VV method will keep V^{BS} hedged in the Black-Scholes world but not in the real market; unless we know the market model perfectly, we cannot guarantee that the barrier option be well hedged in the real market using the $P_{NoTouch}$ -VV method.

Although we do not find a solid underlying reason for adopting $P = P_{ex}$ for the P-VV method, we will explain a possible intuition behind such a choice of P. As we have analysed above, the portfolio in (6.2.3) will be well hedged if we hedge with the VV ω_k up to t_h and set ω_k zero afterwards. Notice the expectation of t_h, $E(t_h) = T_{ex}$ holds. Therefore, on average, in the time period T_{ex} of the total time T, the VV ω_k defined in (5.2.15) are used and in the remaining time of T, the ω_k are set to zero. One might average these two cases to choose the P-VV weights to be $\frac{T_{ex}}{T}\omega_k + \frac{T-T_{ex}}{T}0 = \frac{T_{ex}}{T}\omega_k = P_{ex}\omega_k$. However, the ω_k at different times are different things. The averaging in this way is problematic in theory.

Between the choice of $P = P_{NoTouch}$ and $P = P_{ex}$, we find that the choice of $P = P_{NoTouch}$ has a better underlying reason than the choice of $P = P_{ex}$.

With the P factor, the first issue of the breaking of the hedging argument described early disappears as well. So when the barrier is hit, the P-VV price is correct. Therefore, the P factor at the same time solves both issues, namely the breaking of hedging argument at t_h and the over-hedging at $t < t_h$.

6.2.5. Problems with the P-VV Method. Because $0 \leqslant P \leqslant 1$, the P-VV price is a price between the Black-Scholes price and the unmodified VV price. When the no touch probability is small, the P-VV price is close to the Black-Scholes price. As we have analysed earlier, the P factor is very useful in correcting the VV price, especially when P is small. When the barrier is hit, the P-VV method will correctly price a knockout barrier option at the Black-Scholes price of this option, which is zero. But when the barrier has

not been hit and P is small, the accuracy of the P-VV method is only the same as that of the Black-Scholes method, although the P-VV price may be much better than the VV price in this case. The Black-Scholes price alone is not known for being a very good price for barrier options. Thus the accuracy of the P-VV price may still not be very good when P is small and in the mean time the error of the Black-Scholes price is not insignificant.

The second problem is that the P-VV method does not guarantee that the barrier options priced on the liquid strikes will agree with the market prices of these barrier options. This applies to all exotic options too; notice that if the instrument on the left side of (5.2.23) is not one of those in the right side, the market prices on the liquid strikes will not be produced. Without precisely matching the market prices on the liquid strikes, it is reasonable to think that the P-VV prices between these strikes for barrier options will be less accurate than the VV prices for the vanillas.

There is also a problem with the accuracy of the P estimate. The Black-Scholes implied volatility term structure could affect P_{ex} significantly: if the implied volatility is high at the beginning and low later, P_{ex} will be smaller and if the implied volatility is low at the beginning and high later, P_{ex} will be larger. However, for a $P_{NoTouch}$ estimate made on the Black-Scholes ATM market implied vol, we will show that the accuracy of $P_{NoTouch}$ could affect the P-VV prices but to a very limited extent. The first two problems appear to be the most important factors that affect the accuracy of the P-VV prices.

These problems are related to the fundamental issue that the P-VV method only adds statistically plausible hedging costs to the Black-Scholes price while the VV method adds exact hedging costs to the Black-Scholes price; here the "exact" only means that the hedging costs computed by the VV formula will make the portfolio $dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS}$ well hedged in the Black-Scholes world. In the real world, of course, the hedging will not be exact.

6.3. The V-VV Method - an Alternative or Complement to P

6.3.1. The V-VV Method - Replacing Vanillas U_k with Exotics V_k. In the P-VV method, the barrier option V^{BS} and the vanilla options U_k^{BS} are different types of instruments. We want to explore whether a change from vanilla options U_k^{BS} to exotic options V_k^{BS} in the VV formula has any benefits to pricing. V_k^{BS} are the exotics on the same liquid strikes as for U_k^{BS}; the hedging options V_k^{BS} and the target V^{BS} are of the same

type. There are a few reasons for such a consideration: first, this modified VV method will resemble in form to the highly successful VV method for vanillas because the target option and the hedging options are of the same type; second, we will see that this modified VV method will match the market prices V_k^{mkt} supplied as input, just as in the VV method the market vanilla prices on the liquid strikes can be produced; third, for a down-out barrier option, when the spot hits the barrier, $V_k^{mkt} - V_k^{BS}$ will be zero so this modified VV price will be correctly made equal to the Black-Scholes price V^{BS}. A problem with the modified VV price is that the market prices of the exotics V_k^{mkt} are not market observable. The market prices of the exotics V_k^{mkt} have to be estimated. Hence the V_k^{mkt} estimation error affects the quality of the modified VV price. However, as we will show with numerical examples, this modified VV price with the V_k^{mkt} estimation error in a limited degree may still outperform some other more frequently used methods. We will further discuss the pros and cons after we introduce the definition of this method.

For exotics with a fixed strike parameter, we propose the following modified version of the VV method:

$$V^{V-VV} = V^{BS}\big|_{\sigma_{ATM}} + \sum_{k=1}^{3} \omega_k \left(V_k^{mkt}\Big|_{estimate} - V_k^{BS}\big|_{\sigma_{ATM}} \right) \tag{6.3.1}$$

where k is the index for the exotic option V evaluated at the liquid strike K_k, and ω_k satisfy

$$\begin{aligned}
\frac{\partial V^{BS}}{\partial \sigma}\Big|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \frac{\partial V_k^{BS}}{\partial \sigma}\Big|_{\sigma_{ATM}} \\
\frac{\partial^2 V^{BS}}{\partial \sigma^2}\Big|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \frac{\partial^2 V_k^{BS}}{\partial \sigma^2}\Big|_{\sigma_{ATM}} \\
\frac{\partial^2 V^{BS}}{\partial \sigma \partial S}\Big|_{\sigma_{ATM}} &= \sum_{k=1}^{3} \omega_k \frac{\partial^2 V_k^{BS}}{\partial \sigma \partial S}\Big|_{\sigma_{ATM}}
\end{aligned} \tag{6.3.2}$$

where V_k^{BS} are the same type of instruments as V^{BS}. For example, if the target option V is a knockout barrier option, the V_k will be knockout barrier options on liquid strikes K_k. The target exotic option $V^{BS}\big|_{\sigma_{ATM}}$ is priced at K while V_k are prices at K_k, $k = 1, 2, 3$. On the liquid strikes, only estimated market prices $V_k^{mkt}\big|_{estimate}$ are possible. In this modified form of VV method for barrier options, the vanilla options U_k are replaced by the exotic options V_k of the same maturity at the liquid strikes in the VV formula.

Replacing U with V in (5.2.23) and following the same analysis, we will see that the V^{V-VV} formula is able to produce estimated market prices $V_k^{mkt}\big|_{estimate}$ on the liquid

strikes. If a perfect observation is available on the exotic prices on these liquid strikes so that $V_k^{mkt}\big|_{estimate} = V_k^{mkt}$, the following will hold:

$$V^{V-VV} = V^{BS}\big|_{\sigma_{ATM}} + \sum_{k=1}^{3} \omega_k \left(V_k^{mkt} - V_k^{BS}\big|_{\sigma_{ATM}} \right),$$

and the V^{V-VV} formula will produce the true market prices V_k^{mkt} on the liquid strikes.

Although the market values of the exotic options are not available, the market values at the liquid strikes can be estimated from the P-VV prices (if barrier options are priced) and the Black-Scholes exotic prices at the market Black-Scholes implied vols. Compared to the non-liquid strikes, the liquid strikes have more information for price estimates.

The limitation of this method is that if $V_k^{mkt} - V_k^{mkt}\big|_{estimate}$ are large, this method will result in large errors. This formula will fit the estimated prices $V_k^{mkt}\big|_{estimate}$ exactly on the liquid strikes; therefore, it will not fit the true market prices V_k^{mkt} exactly. To make the method work, the estimated market prices of the barrier options at the liquid strikes must be close to the market prices.

For easy reference to this form of VV method, we will call it the V-VV method.

6.3.2. Applying the V-VV Method to Barrier Options. When the barrier is hit, the V-VV formula will give the correct knock-out barrier price of $V^{V-VV} = V^{BS}\big|_{\sigma_{ATM}} = 0$ because $V_k^{mkt} - V_k^{BS}\big|_{\sigma_{ATM}} = 0$ holds. The V-VV method completely avoids the hedging problem analysed earlier, because the target barrier option V^{BS} and the hedging barriers V_k^{BS} always expire at the same time.

The barrier options at the vol pillars in the V-VV method can be comparable in type to the vanillas in the VV method. But we have some flexibility as to whether a call or put should be used for $V_k^{BS}\big|_{\sigma_{ATM}}$. For instance, if the spot is larger than the barrier and the barrier is larger than K_k, the price and Greeks of the down-out put are zero, making the ω_k difficult to compute. In this case, we can instead use a down-out call for $V_k^{BS}\big|_{\sigma_{ATM}}$ to avoid this difficulty.

Although the market values of the barrier options are not observable on an open exchange, traders do have some information about the prices of the exotic options, for example, the observations of the prices originated from their own dealings with these options. So the V-VV method can be an alternative method to the P-VV method. The V-VV method can work with the P-VV method as its complement too if the P-VV method is used to

compute the barrier prices at three liquid strikes. Then if we have some ways to adjust and improve the P-VV prices based on market factors, we can use them in the V-VV method and let the V-VV method estimate what the prices should be between the liquid strikes.

The V-VV method is proposed only as a secondary method in occasions when several market prices are already estimated. Its mentioning in this thesis is driven by the following reasons:

First, it is interesting to see that the VV method works for non-vanilla products as hedging instruments. This shows that the VV method is more appropriate being viewed as a framework than a model that must theoretically depend on the use of vanilla options in its construction.

Second, it is not a goal to view the V-VV method as a sweeping solution for barrier options in all occasions. But as we will see later, one part of the practical usefulness of the V-VV method that this thesis has identified is this method's input error tolerance for barrier options with small P. It will be demonstrated that with certain not so insignificant amount of input error in estimated barrier option prices, the resulting V-VV interpolation for barrier option with small P is far more accurate than the P-VV and the Black-Scholes prices.

Finally, traders may have much better confidence in estimating prices for exotics on only some strikes because of the pattern of their dealings with the exotics in the market. For example, a trader may have bought and sold barrier options only on some strikes for the past hour; thus it would be much easier for him to make new estimates on those options, given the recent prices he made to or were offered by the counterparties. The V-VV method provides an effortless way to interpolate a trader's estimates for the strikes that lack dealings by the trader.

6.4. Performance under Market Assumptions

6.4.1. Black-Scholes Model as the Market. In this section, we will assume that the flat smile Black-Scholes model actually follows the market but the model makes a wrong estimate about the true market Black-Scholes implied vol. We will compare errors resulting from the VV, $P_{NoTouch}$-VV, P_{ex}-VV, V-VV, Black-Scholes prices of a knock-out barrier call option.

We have studied the error behaviour of the VV method under a flat vol Black-Scholes market assumption in the last chapter. We will now apply the same type of error analysis in the last chapter to compare the P-VV and V-VV methods. The price differences $V^{mkt} - V^{BS}\big|_{\sigma_{ATM}}$ and $U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}$ are in an approximate sense viewed as the Taylor expansion around $V^{BS}\big|_{\sigma_{ATM}}$ and $U_k^{BS}\big|_{\sigma_{ATM}}$.

6.4.1.1. *Error about the P-VV Method.* The Taylor series expansion of $V^{mkt} - V^{BS}\big|_{\sigma_{ATM}}$ and $U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}}$ gives rise to

$$\begin{aligned} & V^{mkt} - \left(V^{BS}\big|_{\sigma_{ATM}} + P\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \right) \\ & = (1-P) \sum_k \omega_k \left. \frac{\partial U_k^{BS}}{\partial \sigma} \right|_{\sigma_{ATM}} \delta\sigma + \frac{1}{2} (1-P) \sum_k \omega_k \left. \frac{\partial^2 U_k^{BS}}{\partial \sigma^2} \right|_{\sigma_{ATM}} (\delta\sigma)^2 \\ & + \frac{1}{6} \left(\left. \frac{\partial^3 V^{BS}}{\partial \sigma^3} \right|_{\sigma_{ATM}} - P \sum_k \omega_k \left. \frac{\partial^3 U_k^{BS}}{\partial \sigma^3} \right|_{\sigma_{ATM}} \right) (\delta\sigma)^3 + O\left((\delta\sigma)^4 \right) \end{aligned} \tag{6.4.1}$$

hence

$$V^{mkt} - \left(V^{BS}\big|_{\sigma_{ATM}} + P\sum_k \omega_k \left(U_k^{mkt} - U_k^{BS}\big|_{\sigma_{ATM}} \right) \right) \approx \begin{cases} O\left((\delta\sigma)^3 \right), & P = 1 \\ O\left(\delta\sigma \right), & P \neq 1 \end{cases} \tag{6.4.2}$$

where $\delta\sigma$ means the Black-Scholes implied vol difference between the market and the Black-Scholes model; the model estimate uses σ_{ATM}. As P approaches 1, the value of the $\delta\sigma$ term will decrease; at some point, the $(\delta\sigma)^3$ term will dominate the error, and the convergence behaviour switches from $O(\delta\sigma)$ to $O\left((\delta\sigma)^3\right)$.

From the above analysis, the factor P cuts the order of error convergence in volatility from cubic to linear. From this view, this does not seem to be a very good method. However, as has been discussed earlier, the factor P corrects the VV pricing error by scaling down the hedging costs. We will show numerical studies on how the P-VV price improves over the VV method in a market with a smile. .

6.4.1.2. *Error about the V-VV Method.* Here we replace the vanillas U_k with barrier options V_k in the VV formula. We obtain

$$\begin{aligned} V^{mkt} - \left(V^{BS}\big|_{\sigma_{ATM}} + \sum_k \omega_k \left(V_k^{mkt}\big|_{estimate} - V_k^{BS}\big|_{\sigma_{ATM}} \right) \right) \approx O\left((\delta\sigma)^3 \right) \\ + \sum_k \omega_k \left(V_k^{mkt} - V_k^{mkt}\big|_{estimate} \right) \end{aligned} \tag{6.4.3}$$

This modified form of the VV method for barrier options has an advantage in price correction power compared to the P corrected Vanna-Volga method if $V_k^{mkt}\big|_{estimate}$ are close estimates of V_k^{mkt}. The error of the modified form is in the third order whereas that of the P corrected form is linear with respect to vol change. Because the market prices V_k^{mkt} are approximated by $V_k^{mkt}\big|_{estimate}$, the total error will include $\sum_k \omega_k \left(V_k^{mkt} - V_k^{mkt}\big|_{estimate}\right)$. So the quality of the estimate $V_k^{mkt}\big|_{estimate}$ is essential in the accuracy of this VV formula.

6.4.1.3. *Error Convergence under a Flat Black-Scholes Market.* The parameters used for testing are $\tau = 1$, $S = 5$, $r_d = 0.03$ and $r_f = 0.02$. The three liquid strikes are $K_1 = 4.49$, $K_2 = 5.32$ and $K_3 = 6.29$. The strike K for the knock-out barrier call option is set to $0.5\,(K_2 + K_3)$. The Black-Scholes implied volatility of the underlying is $\sigma_{mkt} = 0.25$ for the assumed market. The model Black-Scholes implied vol σ_{mod} will be varied to study the convergence behaviours. The Black-Scholes barrier price and Greeks are computed by the closed form formulas. [41] The P factors $P_{NoTouch}$ and P_{ex} are computed by the Black-Scholes model using the σ_{ATM}. The first hitting time probability distribution is given by Wystup [104] and the factors $P_{NoTouch}$ and P_{ex} can be obtained by integrating this first hitting time probability distribution.

The "V-VV Distort" in Figure (6.4.1) denotes the relative error of the estimated market prices at liquid strikes

$$\text{V-VV Distort} = \left(V_k^{mkt} - V_k^{mkt}\Big|_{estimate}\right) / V_k^{mkt}.$$

Figure (6.4.1) shows that the errors of the "V-VV Distort 0%" and VV methods are cubic with respect to $\sigma_{mod} - \sigma_{mkt}$ changes. "V-VV Distort 0%" means V-VV Distort = 0%. The factor $P_{NoTouch}$ or P_{ex} in the P-VV method decreases the order of error convergence from cubic to first order.

The errors of the V-VV prices with market price estimate errors on liquid strikes are dominated by the market price estimate errors, hence demonstrating a flat error.

According to the P-VV formula, both the $P_{Notouch}$-VV and P_{ex} -VV prices are always in between the Black-Scholes price and the VV price. Because P_{ex} is larger than $P_{Notouch}$, the P_{ex} -VV price is always closer to the VV price than the $P_{Notouch}$-VV price; the $P_{Notouch}$-VV price is always closer to the Black-Scholes price than the P_{ex} -VV price. This observation can be clearly seen in Figure (6.4.1). As the barrier is closer to the spot in the bottom plot of Figure (6.4.1), both the $P_{Notouch}$-VV and P_{ex} -VV prices will be closer to the

Black-Scholes price; otherwise, they will be closer to the VV price as shown in the top plot.

Figure (6.4.2) gives the error behaviours of the P-VV price under different P values. Except the value of P, all the parameters are kept the same in Figure (6.4.2) as those in the top plot of Figure 6.4.1. The equality (6.4.1) suggests that the P-VV price error under a flat vol Black-Scholes market should see a switch from the linear to the cubic error convergence as P tends to 1. Figure (6.4.2) shows that this switch of error behaviour happens when P is close to 1.

Following Elder [32], we will then move to a more realistic market assumption by setting the Heston model as the market.

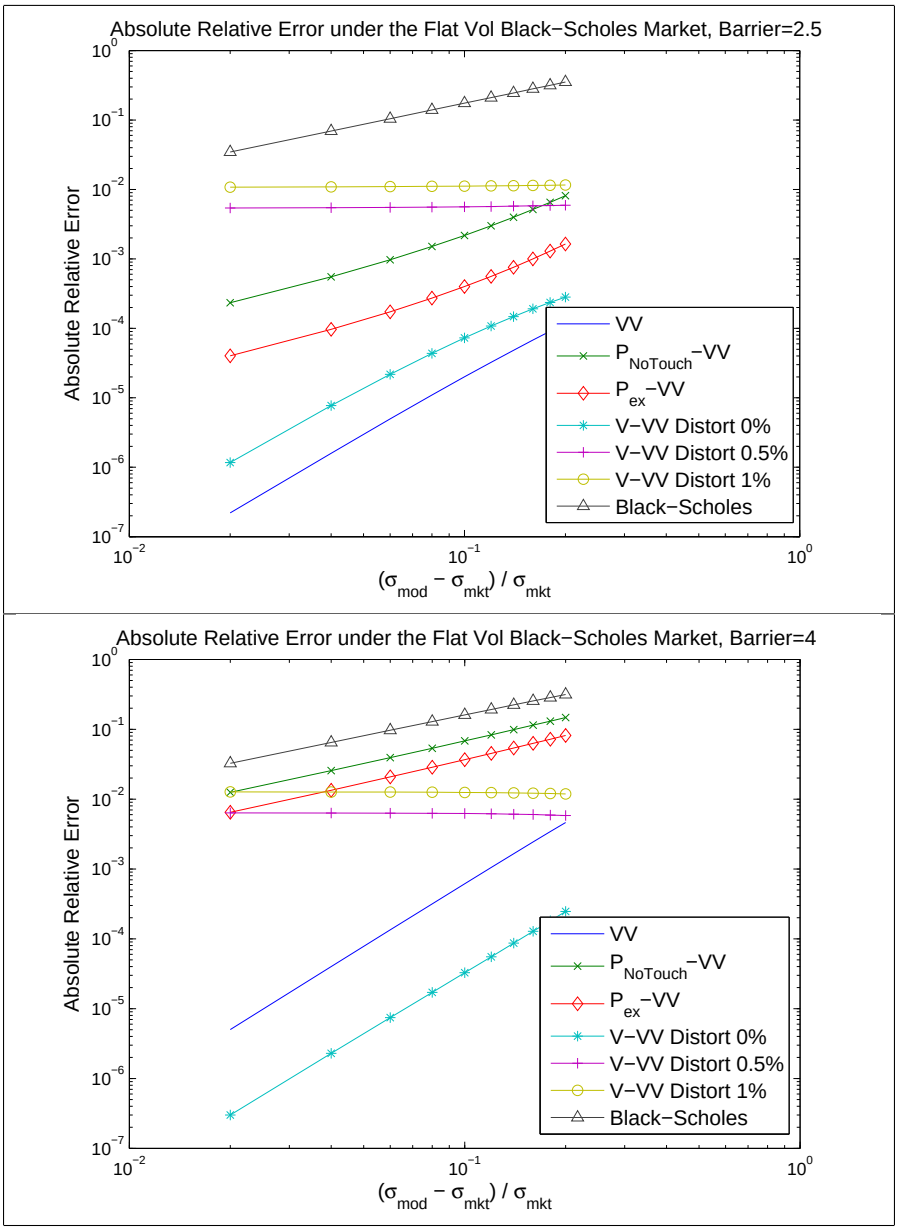

FIGURE 6.4.1. Order of Convergence with Barrier= 2.5 (top) and Barrier= 4 (bottom)

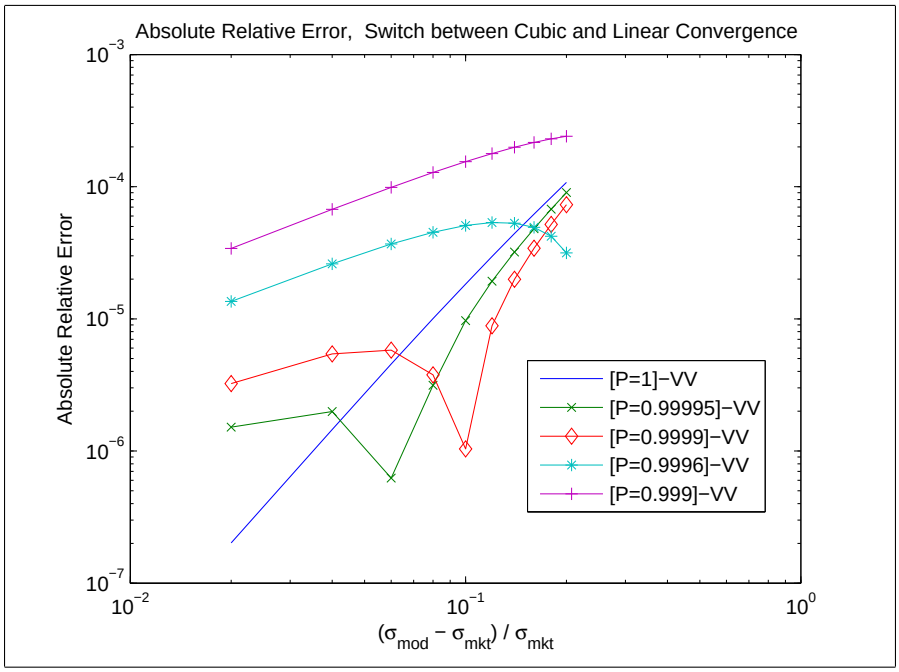

FIGURE 6.4.2. Order of Convergence Changes between Cubic and Linear with the P Value, Barrier= 2.5

6.4.2. Heston Model as the Market. The parameters for the Heston model and the option are $\kappa = 1.1$, $\theta = 0.09$ $\xi = 0.27$, $\rho = -0.7$, $\nu_0 = 0.09$, $\tau = 1$, $S = 5$, $r_d = 0.03$ and $r_f = 0.02$. The parameters for three liquid strikes are $K_1 = 4.3$, $K_2 = 5.05$ and $K_3 = 5.7$. The target option is a down and out barrier call option. The Heston market barrier prices are computed by the finite difference method with the Crank-Nicolson, ADI and Craig-Sneyd schemes implemented. The Black-Scholes price and Greeks are computed with the closed form formulas [41]. The P factors $P_{NoTouch}$ and P_{ex} are computed by the Black-Scholes model using the σ_{ATM}. Figures (6.4.3), (6.4.4) and (6.4.5) show the relative error comparison of the VV, $P_{NoTouch}$-VV, P_{ex} -VV, Smile-VV, V-VV and Black-Scholes prices under different no touch probabilities.

Figure (6.4.3) shows the case of the spot being far away from the barrier and the no touch probability being high. In this case, the P-VV prices are close to the VV price and the barrier option is close to a vanilla option. Conceptually, the Smile-VV method is the same as the VV method for pricing a vanilla option. So in this case, the Smile-VV price is also close to the VV price. Because the VV method works very well for vanilla options as we have seen in the last chapter, the P-VV, Smile-VV and VV prices all show excellent relative error in the order of 10^{-3}. Relatively speaking, the VV price works better than the P-VV prices; between the two choices of P-VV prices, the P_{ex} -VV price is better than

the $P_{NoTouch}$-VV price because it is closer to the VV price. However, they all seem to be very good prices. The V-VV price is the best price if the $\left(V_k^{mkt} - V_k^{mkt}\big|_{estimate}\right)/V_k^{mkt}$ is zero; with 0.5% or 1% biases in the $V_k^{mkt}\big|_{estimate}$, the V-VV price is not as good as the P-VV and Smile-VV prices.

Figure (6.4.4) shows the case of the no touch probability being in the middle range of [0, 1]. In this case, the P-VV and Smile-VV prices have less than half the error of the VV price; in this particular test, both the P-VV and Smile-VV prices are around 2×10^{-2}. Now the V-VV price with a bias $\left(V_k^{mkt} - V_k^{mkt}\big|_{estimate}\right)/V_k^{mkt}$ up to 1% appears better than the P-VV and Smile-VV prices.

Figure (6.4.5) shows the case of the no touch probability being small. According to our analysis earlier, the P factor should be particularly useful in correcting the VV price for this case. As we can see on Figure (6.4.5), if uncorrected, the VV price behaves very badly; the relative error even goes over 40% for some strikes whereas the $P_{NoTouch}$-VV and P_{ex} -VV prices have the errors consistently within 6% across the displayed range. From Figures (6.4.5) and (6.4.4), we see that as an overall trend, the correcting power of the P factor becomes larger as the no touch probability becomes smaller. When the no touch probability is small, the P-VV price is closer the Black-Scholes price, which appears to be a good price now. Figure (6.4.5) shows that the Smile-VV price is still a very good price in this case.

Overall, the Smile-VV price keeps a relatively flat error behaviour at each of these cases of no touch probability.

The difference between the two choices of P factors is small; this may explain why both are used in practice. However, unless the no touch probability is very large, these figures suggest that the $P_{NoTouch}$-VV price is a better choice overall than the P_{ex} -VV price. Even when the no touch probability is very large, the small error of the $P_{NoTouch}$-VV price makes it a very good price too. What is also conspicuous is that the Smile-VV appears to be a price roughly as good as the $P_{NoTouch}$-VV price for all three cases of no touch probability. In Figure (6.4.4), the errors of the $P_{NoTouch}$-VV price and the Smile-VV price are in opposite signs, suggesting the average of the two prices may lower the error in some occasions.

The V-VV method is not a primary method because it cannot be used alone; in order for it to work, it has to be combined with some liquid strike market barrier option price

estimation method. If the market barrier option prices are wrongly estimated, the error of the V-VV price can be large. However, it can be very accurate if the estimates $V_k^{mkt}|_{estimate}$ are very accurate.

We have tested using the Black-Scholes implied vol from the VV vanilla price to compute no touch probability; we have also tested on estimating the no touch probability with the Heston model using Monte Carlo. No significant shifts of error behaviours have been observed. The P estimation error using σ_{ATM} seems to be minor in an ordinary market parameter settings. An example is shown in Figure (6.4.6), for which the computation for the $P_{NoTouch}$-VV and P_{ex} -VV prices is the same as in Figure (6.4.4) except that in Figure (6.4.6), $P_{NoTouch}$ and P_{ex} are computed using not σ_{ATM} but the Black-Scholes implied vol of the VV vanilla price at each K, denoted by $\sigma^{VV}(K)$. The effects are barely noticeable in this particular example. But it should be cautioned that in some extreme market conditions when the estimation using the ATM vol might not suffice, an estimate of the P factor using the ATM Black-Scholes implied vol could increase the error of the P-VV price to a noticeable level.

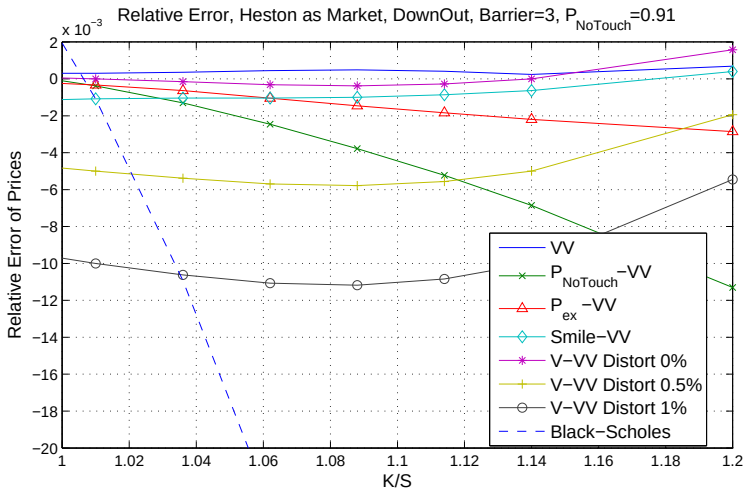

FIGURE 6.4.3. Relative Error for Down-Out Call, Heston as Market, $P_{NoTouch} = 0.91$

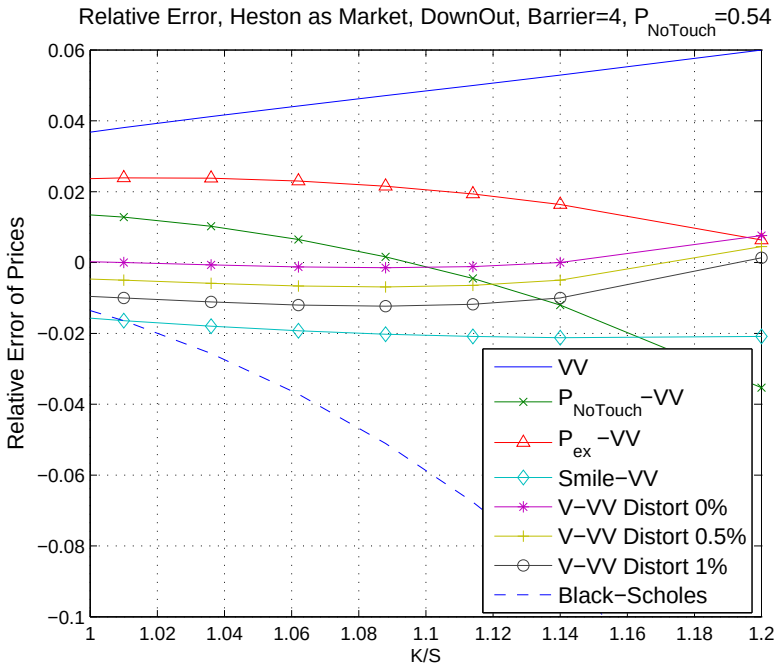

FIGURE 6.4.4. Relative Error for Down-Out Call, Heston as Market, $P_{NoTouch} = 0.54$

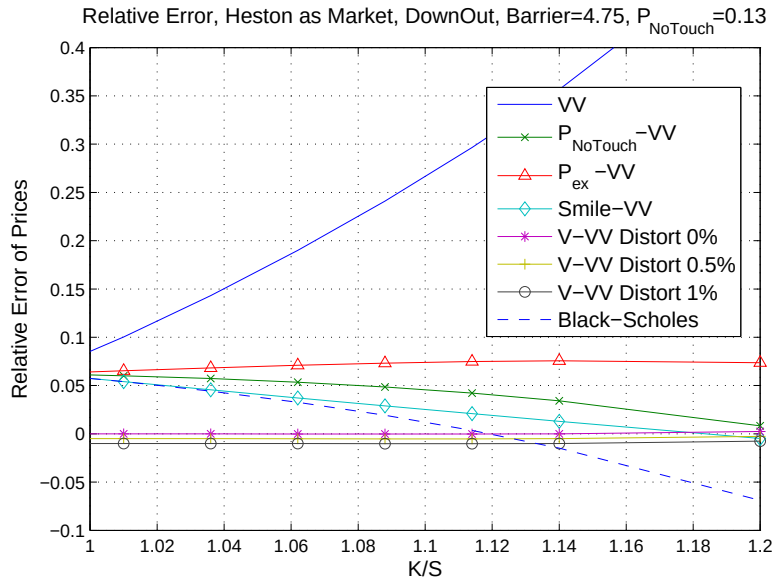

FIGURE 6.4.5. Relative Error for Down-Out Call, Heston as Market, $P_{NoTouch} = 0.13$

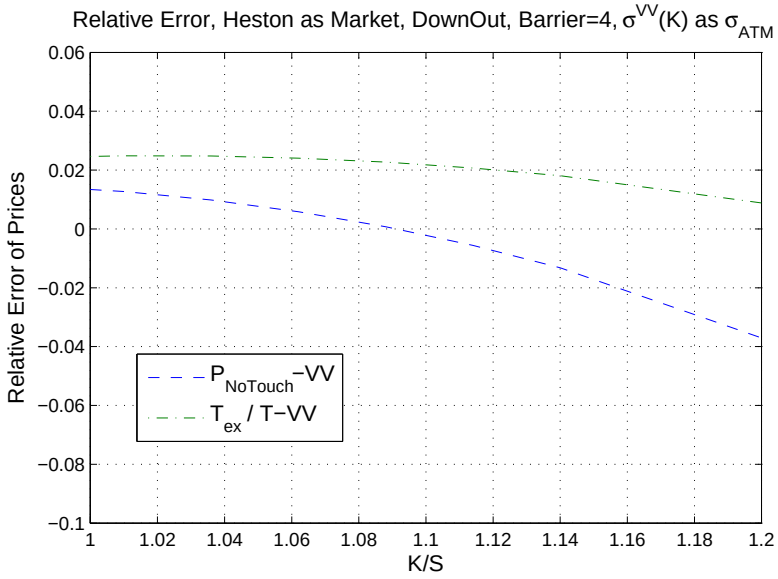

FIGURE 6.4.6. Relative Error for Down-Out Call, Heston as Market, Barrier= 4, ($P_{NoTouch}$ and P_{ex} Computed by the VV interpolated volatility $\sigma^{VV}(K)$ Instead of σ_{ATM})

6.4.2.1. *Error with respect to the Barrier.* Figure (6.4.7) shows the error between the Heston market price and the other prices as listed in the plot. The same parameters as in the last section are assumed except that the strike K is fixed and the barrier level now changes; the strike K is set to be $0.5\,(K_{atm} + K_{25c})$ in the top plot and to $1.2S$ in the bottom plot. The strike $K = 1.2S$ is in the extrapolation range outside of the liquid strike K_{25c}.

Without the P factor, the VV price will not be zero when the spot hits the barrier. As the barrier moves closer to the spot, the $P_{NoTouch}$-VV and P_{ex}-VV prices show significant improvement over the unmodified VV price. The error of the Black-Scholes price improves as the barrier is closer to the spot; at a point very close to the spot, the Black-Scholes price becomes better than the VV and P-VV prices.

The $P_{NoTouch}$-VV price performs better than the P_{ex}-VV price when the strike is less out-of-the-money as shown in the top plot; the P_{ex}-VV price outperforms the the $P_{NoTouch}$-VV price around barrier being $0.5S$ to $0.8S$ when the strike is more out-of-the-money as shown in the bottom plot. The $P_{NoTouch}$-VV price seems to be better when the strike is between the liquid strikes, hence good at interpolation; as the strike goes far out of the money, the P_{ex}-VV price starts to show better performance, especially in the extrapolation range of the strike.

The V-VV prices are better than the other prices shown in Figure (6.4.7). In the extrapolation range of the strike, the performance of the V-VV prices are much better than the other prices shown in Figure (6.4.7). Of course, this satisfying performance must be achieved by accurate estimates about $V_k^{mkt}\big|_{estimate}$. If the estimates about $V_k^{mkt}\big|_{estimate}$ carry large errors, the V-VV prices will have large errors as well.

The Smile-VV price overall performs better than the P-VV prices, demonstrating a rather flat error curve with respect to the barrier. The interpolation and extrapolation performances are both better compared to the P-VV prices, suggesting the Smile-VV price may be a sound alternative to the P-VV prices too.

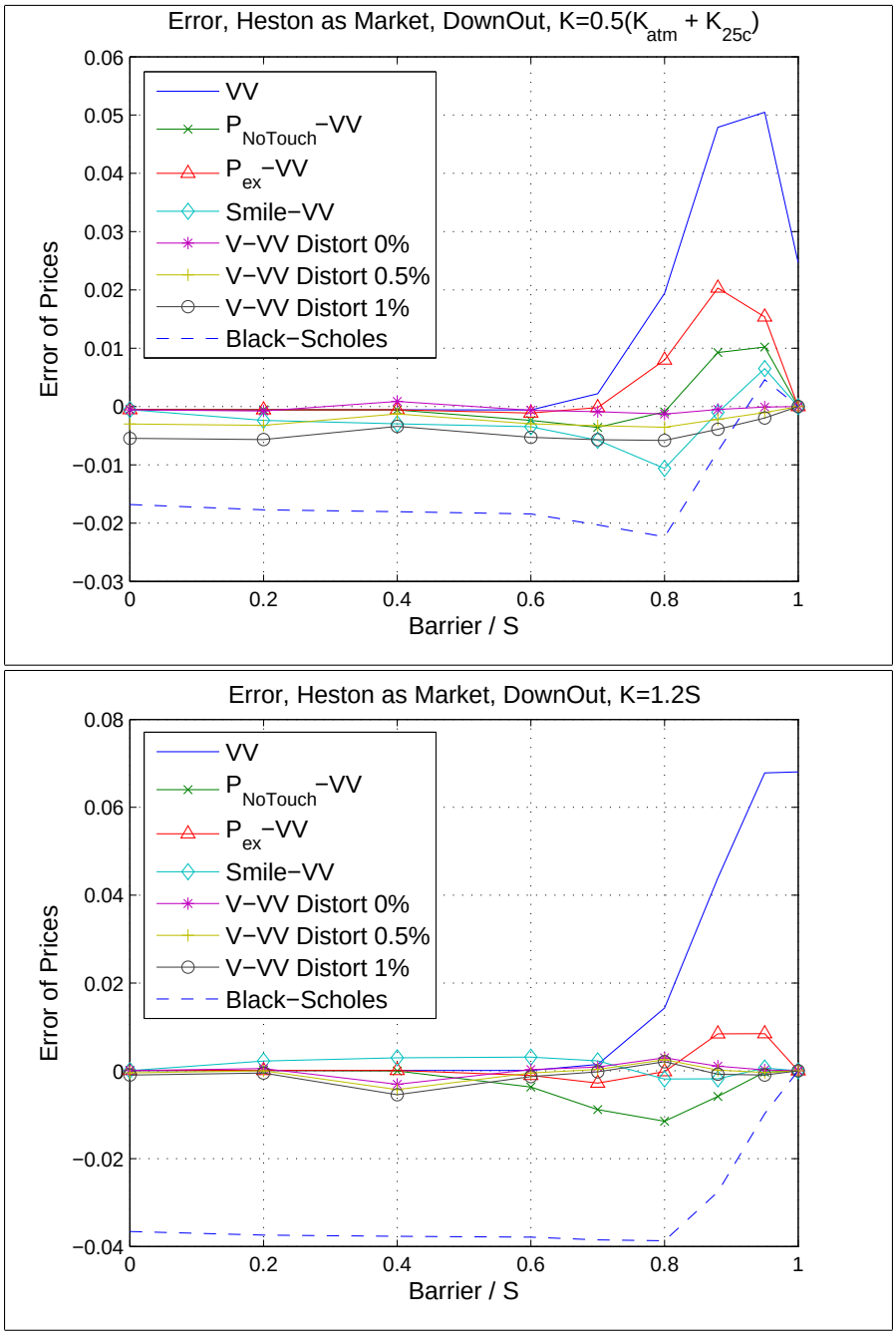

FIGURE 6.4.7. Error for Down-Out Call with respect to the Barrier for $K = 0.5\,(K_{atm} + K_{25c})$ (Top) and $K = 1.2S$ (Bottom)

6.4.2.2. *Error Affected by the Choice of Model Vol Level.* In the VV and P-VV methods, σ_{ATM} is usually chosen to be the Black-Scholes model implied volatility. We will explore whether a change in the model implied volatility can influence the pricing error. Figure (6.4.8) gives a comparison between different Black-Scholes model implied vol choices for the $P_{NoTouch}$-VV method. Except the different Black-Scholes model implied vol choices, all other parameters stay the same as those for Figure (6.4.4).

Despite having a slightly higher error than the other two choices at some strikes, the Black-Scholes implied vol of the VV vanilla price as denoted by $\sigma^{VV}(K)$ appears to have a small overall error across a wide range of strikes. The Black-Scholes barrier option price computed using $\sigma^{VV}(K)$ is the Smile-VV price. Therefore, the $P_{NoTouch}$-VV price using $\sigma^{VV}(K)$ is to add a price correction to the Smile-VV price, which is already an accurate price; on the other hand, $P_{NoTouch}$-VV price using σ_{ATM} is to add a price correction to the Black-Scholes price on the ATM vol, which is a less accurate price than the Smile-VV price. So the $P_{NoTouch}$-VV price using $\sigma^{VV}(K)$ seems to enjoy a better starting point. The $P_{NoTouch}$-VV price using $(\sigma_{25p} + \sigma_{ATM} + \sigma_{25c})/3$ as the model vol has some improvement over the $P_{NoTouch}$-VV price using σ_{ATM} as the model vol roughly after $K/S = 1$. For convenience, we define $\sigma_{AVG} = (\sigma_{25p} + \sigma_{ATM} + \sigma_{25c})/3$. An average of all three market Black-Scholes implied volatilities is possible to reduce error because the resulting average volatility differences between the model vol and the three liquid vols may be smaller, hence making the overall price error smaller. As a rule of thumb we can suggest based on our numerical examples, a sound alternative to σ_{ATM} is the average of the three market Black-Scholes implied vols $(\sigma_{25p} + \sigma_{ATM} + \sigma_{25c})/3$. The choice $\sigma^{VV}(K)$ may also be used and its extrapolation seems to be better. Other choices have not been found to benefit the error behaviour of the P-VV method.

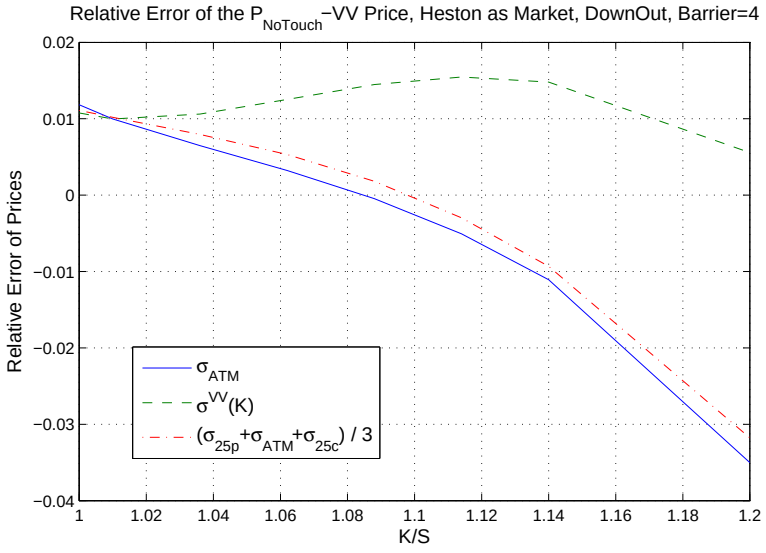

FIGURE 6.4.8. Relative Error for Down-Out Call under Different Model Vol Choices, Heston as Market

6.4.2.3. *Comparing the V-VV Method Applied to Other Path Dependent Exotics.* The V-VV method gives a sound interpolation method if the barrier options on the liquid strikes

are accurately priced. To other path dependent options such as Asian and lookback options, the V-VV method will not perform as well as it is for the barrier options. According to the analysis in the last chapter, the VV ω_k will not make all the Vanna terms balanced in the portfolio $dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS}$ for some path dependent options such as Asian and lookback options. By similar analysis, the V-VV ω_k defined in (6.3.2) will not make all the Vanna terms balanced in the portfolio $dV^{BS} - \Delta dS - \sum_k \omega_k dV_k^{BS}$ for Asian and lookback options either because the terms such as $\frac{\partial V^{BS}}{\partial \sigma \partial \overline{S}}$, $\frac{\partial V_k^{BS}}{\partial \sigma \partial \overline{S}}$ and etc will not be cancelled out just by the V-VV weights ω_k. However, since V_k^{BS} are V^{BS} are the same type of instruments in the V-VV method, the unbalanced stochastic components in $dV^{BS} - \Delta dS - \sum_k \omega_k dV_k^{BS}$ may be smaller than those in the VV method.

Figure (6.4.9) shows the VV price relative error comparison among a down and out barrier call, an arithmetic Asian call, and a fixed strike lookback call prices under a Heston market assumption. All of the Heston and option parameters are kept the same as those used for the computation in the last section; that means that all three options have the same strike ($K = 0.5\,(K_2 + K_3)$); the maturity and the spot are the same for all of them; the barrier option has the barrier equal to 4. The same computational methods as discussed in the last section are used to compute the Heston market prices and Black-Scholes model prices for all these options. The VV method is assumed to have zero bias: $\left(V_k^{mkt} - V_k^{mkt}\big|_{estimate}\right) = 0$ in the market price estimates.

Figure (6.4.9) shows the errors for Asian and lookback options are generally larger than those for barrier options. Note all three plots of Figure (6.4.9) have equal ranges of ξ and ρ; the mesh plots are presented in angles from which the surfaces can be best viewed. The top plot of Figure (6.4.9) shows that the relative error is well within 1%; the middle plot and the bottom plot both have errors over 1% for some parameters. This result agrees with the early analysis about the unbalanced Vanna terms of Asian and lookback options.

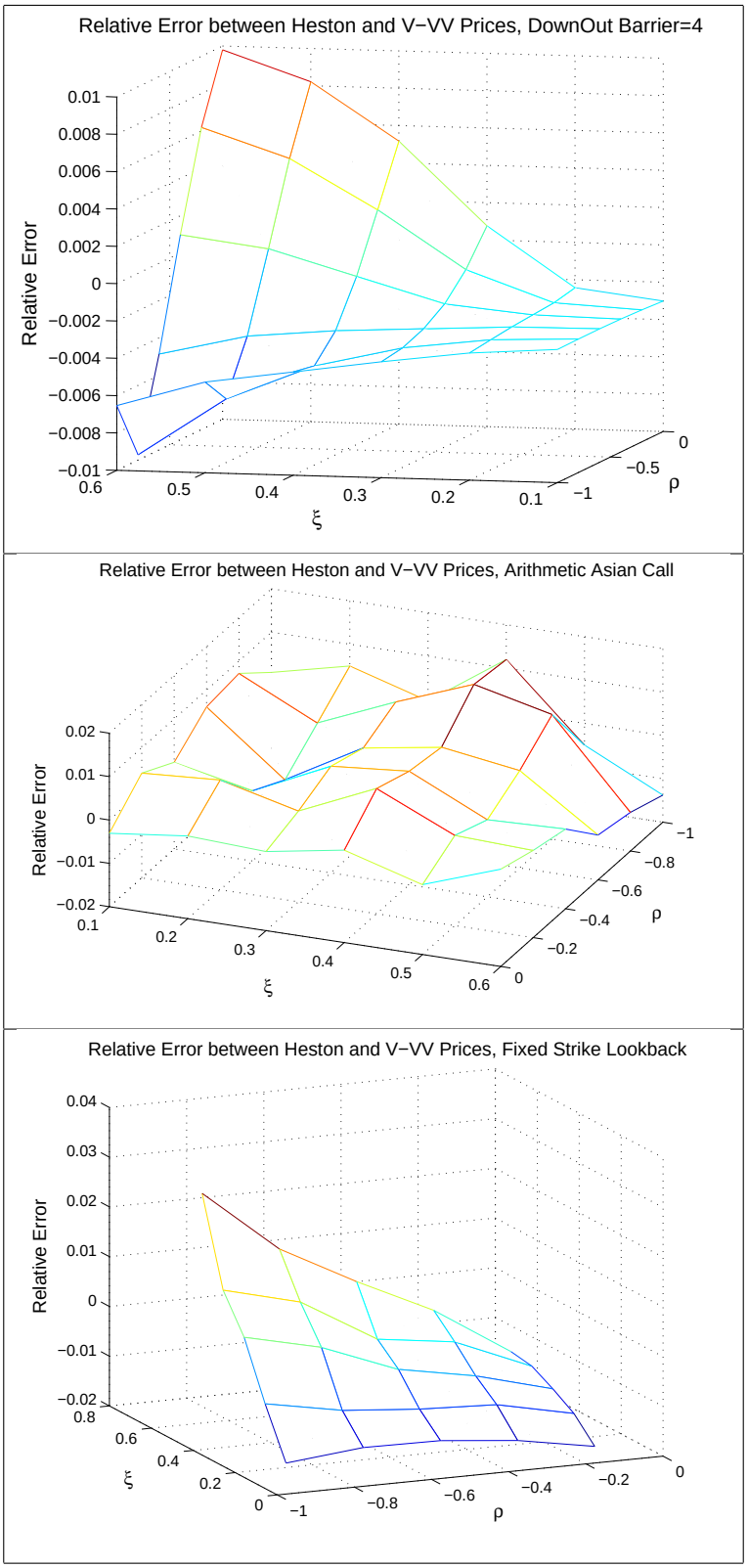

FIGURE 6.4.9. Relative Error for Down-out Barrier Call (Top), Arithmetic Asian Call (Middle) and Fixed Strike Lookback Call (Bottom), Heston as Market

6.4.2.4. *Comparison of the P-VV price and the Smile-VV price under Different Market Parameters.* In this section, we will study the performance of the P-VV and Smile-VV prices under different market smiles by changing the assumed market Heston model parameters.

We will now focus on a comparison between the P-VV and Smile-VV methods. The errors of the P-VV and Smile-VV prices seem to be roughly of similar sizes as in the early discussion on the relative error changing with respect to strike. The P-VV method is used in banking practice but the Smile-VV method is not. So it is interesting to see whether the Smile-VV price is as good as the P-VV prices for a wide range of market parameters.

All the parameters for Figures (6.4.10), (6.4.11), (6.4.12) are set the same as those used for the barrier option computation in the last section. The plots on the left side have the strike $K = K_2$ and the plots on the right side have the strike $K = 0.5\,(K_2 + K_3)$. These figures suggest that the Smile-VV is roughly as accurate as the P-VV method. From these figures, we learn that the P_{ex}-VV and Smile-VV methods perform better than the $P_{NoTouch}$-VV method when the strike K is further out of the money.

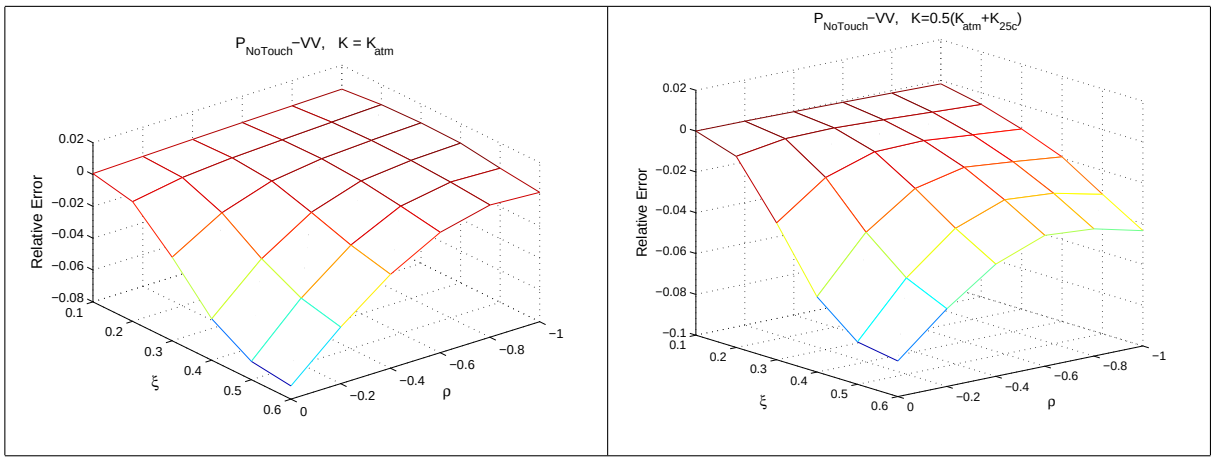

FIGURE 6.4.10. Relative Error of the $P_{NoTouch}$-VV Price for Down-Out Call, Heston as Market, Barrier= 4

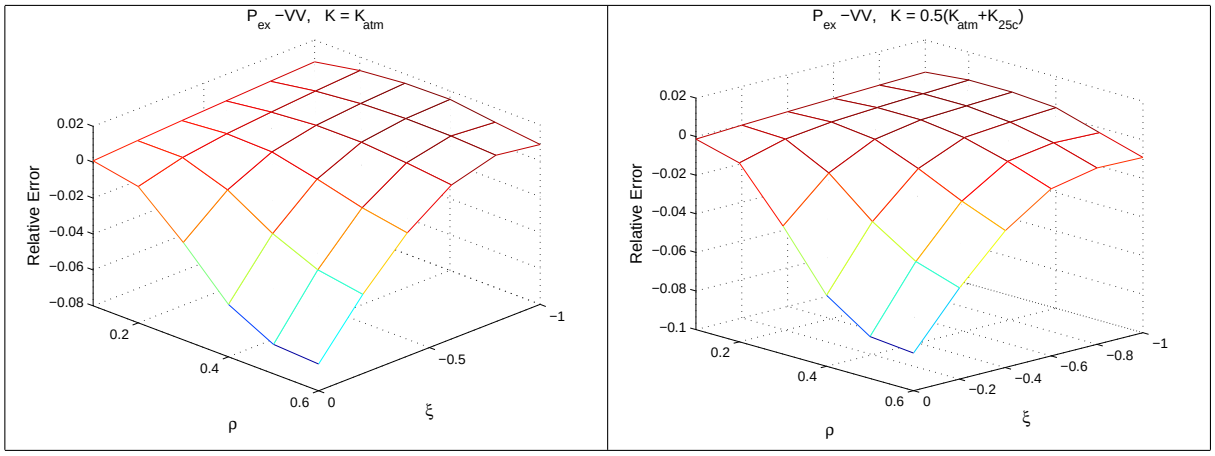

FIGURE 6.4.11. Relative Error of the P_{ex}-VV Price for Down-Out Call, Heston as Market, Barrier= 4

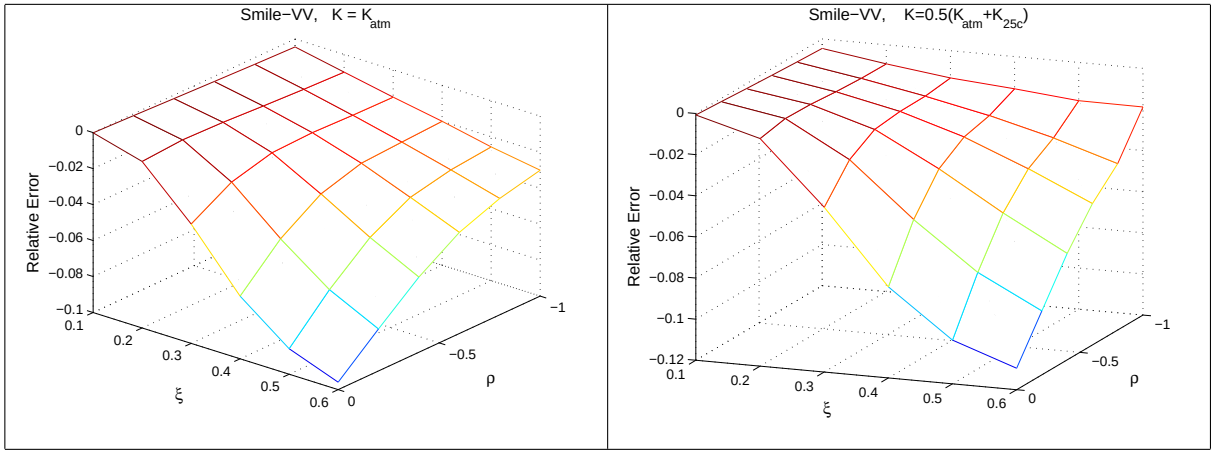

FIGURE 6.4.12. Relative Error of the Smile-VV Price for Down-Out Call, Heston as Market, Barrier= 4

6.5. Summary of Key Findings

(1) The P factor in the P-VV method is related to a correction to the hedging costs added to the Black-Scholes price; the VV hedging weights ω_k alone will not be enough for the barrier options. Two issues with the unmodified VV method are identified for the down-out barrier options: the inconsistency of the VV price at t_h and over-hedging at $t < t_h$. Both P factors solve the first issue completely. From a view of solving the second issue, the $P_{NoTouch}$-VV method is found to have better underlying reason than the P_{ex}-VV method. (2) When the no touch probability is small, the $P_{NoTouch}$-VV and P_{ex}-VV prices are much better than the VV price; this suggests a divergence of the VV price as the no touch probability becomes smaller. The error of the P-VV price is close to the error of the Black-Scholes price when the no touch probability is small. Therefore, although the P-VV

price may be a big improvement over the VV price when the no touch probability is small, any further improvement on error is limited by how close the Black-Scholes price is to the market price in this case. (3) If the market barrier option prices on the liquid strikes can be accurately estimated, the V-VV method can help to interpolate prices between these strikes with relatively high accuracy. The accuracy of the V-VV price relies on the accuracy of the estimates of $V_k^{mkt}\big|_{estimate}$. (4) The V-VV method works better for barrier options than for Asian and lookback options because there is still an unbalanced Vanna problem for Asian and lookback options. (5) The estimation of $P_{NoTouch}$ and P_{ex} with the Black-Scholes model using σ_{ATM} is good enough for normal market parameters. (6) $\sigma^{VV}(K)$ and σ_{AVG} are useful alternatives to σ_{ATM} as the model vol for the P-VV price computation. (7) The errors between the Heston price and the $P_{NoTouch}$-VV, P_{ex}-VV, Smile-VV prices are of the similar order for down-out barrier options. Within the same order, each method may outperform another depending on the strike and other parameters. The overall trend is that in the interpolation range of the strike, the $P_{NoTouch}$-VV price usually outperforms the P_{ex}-VV price; in the extrapolation range of the strike, the P_{ex}-VV price appears to have an advantage over the $P_{NoTouch}$-VV price. The relative error of the Smile-VV price is relatively flat with respect to the strike and the barrier level compared to the P-VV method; the Smile-VV price appears to be a very good price in both the interpolation and the extrapolation ranges of the strike. This suggests that the Smile-VV price may be used as a sound alternative to the P-VV prices.

CHAPTER 7

The Karasinski Method

7.1. A Generalisation of the VV Method

The VV method provides a fast pricing tool that is based on the flat-vol Black-Scholes model. We have shown in the previous chapters that the VV method is an excellent alternative to the Heston, SABR, CEV and Merton Jump models. This method is very fast because it requires no model calibration at all.

But if the model used in the VV method is not limited to the flat vol Black-Scholes model, will the information on the market vanilla prices improve the mark-to-market price? When the model is not well calibrated, will the market vanilla prices help to correct this calibration error? Karasinski [55] first raised this question and a special case of the Karasinski generalisation (the general case to be discussed shortly) follows

$$V^{Kar} = V^{mod}\Big|_{x^{mod}} + \sum_k \omega_k \left(U_k^{mkt} - U_k^{mod}\Big|_{x^{mod}} \right) \tag{7.1.1}$$

and the weights $\omega_k,\ 1 \leqslant k \leqslant L$ (L is the total number of liquid vanillas; see below) are chosen to satisfy

$$\begin{aligned} \frac{\partial V^{mod}}{\partial x}\Big|_{x^{mod}} &= \sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x}\Big|_{x^{mod}} \\ \frac{\partial^2 V^{mod}}{\partial x^2}\Big|_{x^{mod}} &= \sum_k \omega_k \frac{\partial^2 U_k^{mod}}{\partial x^2}\Big|_{x^{mod}} \\ \frac{\partial^2 V^{mod}}{\partial x \partial S}\Big|_{x^{mod}} &= \sum_k \omega_k \frac{\partial^2 U_k^{mod}}{\partial x \partial S}\Big|_{x^{mod}} \end{aligned} \tag{7.1.2}$$

where "Kar" is an abbreviation of Karasinski, "mod" means model, x denotes a model parameter, x^{mod} is the model parameter value currently used and it may be slightly wrong compared to the value if the model were well calibrated to the market, $V^{mod}\big|_{x^{mod}}$ is the model target option price, $U_k^{mod}\big|_{x^{mod}}$ denote the model liquid vanilla prices, $U_k^{mod}\big|_{x^{mod}}$ denote the market liquid vanilla prices.

The model parameter x moves stochastically with time. The value of x used by the model Greeks is different from the true value of the market; the model is not restricted to the Black-Scholes model and can be a model which already takes into account the vol

smile. For the Black-Scholes model, x means the Black-Scholes implied volatility. If the model is the CEV model, x may be the parameter σ in (5.5.3) for example.

Depending on the market, the number of vanilla instruments L is not necessarily limited to three. If $L = 3$, the Karasinski weights ω_k have unique solutions. If $L > 3$, Karasinski recommends that ω_k are chosen so $\sum_k \omega_k^2$ is minimised. The case of $L < 3$ is not realistic for a market.

In Karasinski's original work, the parameter x was written as σ. In many smile-generating models, there is no single volatility parameter. So we use x here to indicate that it is just one parameter of the model just to avoid confusion. In an extended form of the Karasinski method, the parameter x can be a vector of parameters $(x_1, x_2...x_N)$, which we will later discuss. We will specify the meaning of the model parameters x (depending on the model choice) in each case of our discussion later.

Karasinski does not provide a name for this model. For easy reference, we will refer to this model as the Karasinski method. It should be noted that the Karasinski generalisation defined in (7.1.1) and (7.1.2) is only a special case of the general Karasinski method. The general Karasinski method is a broad framework that consists of a model price and a correction part. There may very well be other ways to compute the correction part other than using the balance of sensitivities w.r.t. model parameters. Because this special case of the Karasinski method, which relies on the balance of sensitivities w.r.t. model parameters, offers sound computational convenience, we will limit the discussion of the Karasinski method to this special case only in this thesis.

We call this special case of the Karasinski method the VVV-Karasinski method; VVV stands for Vega, Volga and Vanna. It should be noted that all of the methods and variants to be introduced later are part of the Karasinski method in general. For easy reference, we will simply call the VVV-Karasinski method the Karasinski method.

We can see from the illustration in Figure (7.1.1) that the Karasinski method becomes the VV method when $V^{mod}\big|_{x^{mod}}$ and $U_k^{mod}\big|_{x^{mod}}$ follow the flat vol Black-Scholes model and x^{mod} is the model Black-Scholes implied vol. In the first plot, the Black-Scholes method only recognises a flat Black-Scholes implied vol curve whereas in the second, the model in the Karasinski method can be priced on a non-flat vol curve with a single parameter x.

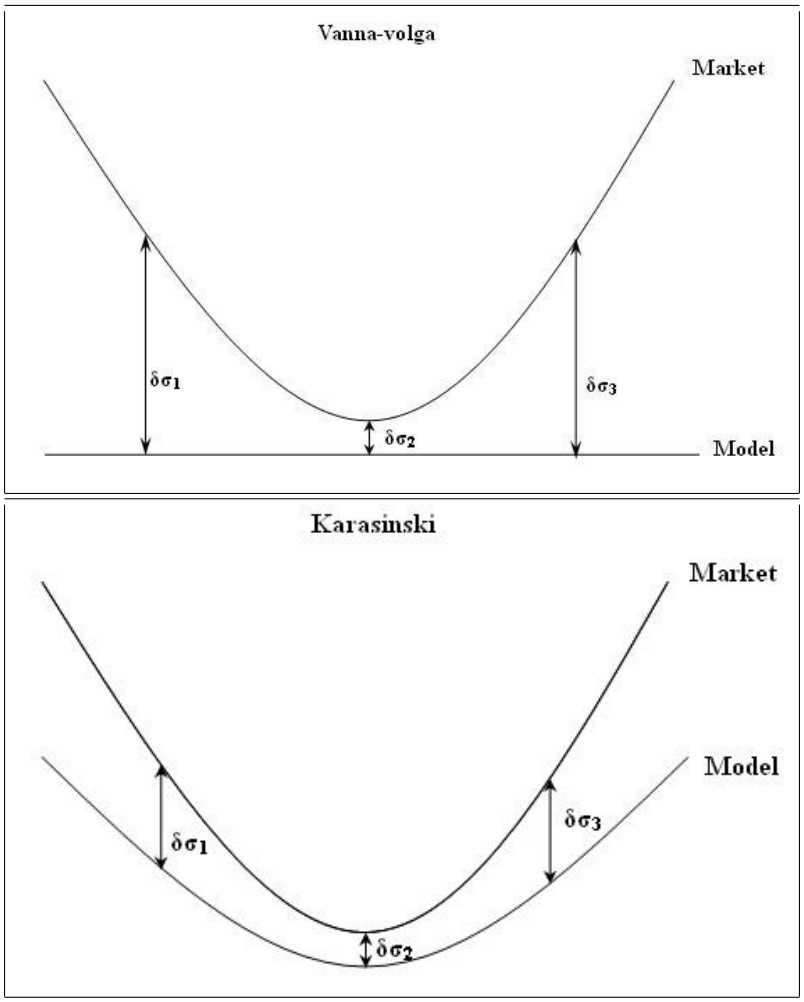

FIGURE 7.1.1. Comparison between Vanna-Volga and Karasinski Methods (σ is the Black-Scholes implied volatility.)

7.1.1. The Hedging-Costs View and the Parameter-Error View. In what we call the parameter-error view of the Karasinski method, the market price is the model price with correct model parameter values, and the model price is the model price with wrong model parameter values due to imperfect calibration. Assuming the Taylor expansions exist, the market-model price differences of the target option V and the vanillas U_k can be expressed by

$$\begin{aligned} V^{mkt} - V^{mod} &= \left.\frac{\partial V^{mod}}{\partial x}\right|_{x^{mod}} \delta x + \frac{1}{2} \left.\frac{\partial^2 V^{mod}}{\partial x^2}\right|_{x^{mod}} (\delta x)^2 + O\left((\delta x)^3\right) \\ U_k^{mkt} - U_k^{mod} &= \left.\frac{\partial U_k^{mod}}{\partial x}\right|_{x^{mod}} \delta x + \frac{1}{2} \left.\frac{\partial^2 U_k^{mod}}{\partial x^2}\right|_{x^{mod}} (\delta x)^2 + O\left((\delta x)^3\right) \end{aligned} \tag{7.1.3}$$

Because the market-model spot difference δS is zero as S is observable, it is not included in the above expansions. In practice, there could be minor factors that make δS nonzero.

For example, there may be a slight time gap between making the fresh observations of U_k^{mod} and making the Karasinski computation using S. Because S is a market observable value, δS can be considered insignificant. In the parameter-error view of the Karasinski method, it is clear to see that the balance of Vanna terms

$$\left.\frac{\partial^2 V^{mod}}{\partial x \partial S}\right|_{x^{mod}} = \sum_k \omega_k \left.\frac{\partial^2 U_k^{mod}}{\partial x \partial S}\right|_{x^{mod}} \tag{7.1.4}$$

is not needed in (7.1.2).

In the hedging-costs view of the Karasinski, on the contrary, we will see that the balance of Vanna terms is necessary. Just like σ_{ATM} in the VV method, assume the parameter x follows a function of diffusion processes. If the target option $V^{mod}\big|_{x^{mod}}$ and the vanillas $U_k^{mod}\big|_{x^{mod}}$ depend on no other stochastic variable but S and x, following the analysis for the VV method, we can easily verify that the Karasinski weights (7.1.2) will make the following equality hold:

$$\begin{aligned} & dV^{mod} - \Delta dS - \sum_k \omega_k dU_k^{mod} \\ & = r_d \left[V^{mod} - S\Delta - \sum_k \omega_k U_k^{mod} \right] dt + r_f S \Delta dt \end{aligned} \tag{7.1.5}$$

As we made the same statement for the growth of $dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS}$ for the VV method, it should be noted that the risk-free growth of the portfolio in (7.1.5) for the Karasinski method is only true in the model world; it does not mean that the real market portfolio value $dV^{mkt} - \Delta dS - \sum_k \omega_k dU_k^{mkt}$ grows at the risk-free rate. Without (7.1.4), the equality (7.1.5) will not hold. For the VV method where *mod* means the Black-Scholes model, we have shown that the balance of the Vanna terms is essential; without it, large errors between liquid strikes may appear (e.g. see Figure (5.5.9)).

So the question is which view is more appropriate. We find that both views can be correct, depending on the market-model relationship.

In the case of the model not following the market, such as the VV method (a special case of the Karasinski method), the hedging-costs view is more appropriate. Because the model and the market are very different, to build a link between V^{mkt} and V^{mod} needs to rely on hedging. The hedging costs are computed in the way that ensures that the portfolio in (7.1.5) grows at rate r_d in the model world. The estimated market price is then based on V^{mod} plus the hedging costs. The Vanna term in this view is important because it

gives the Vega shift due to the spot movement. With the balance of the Vanna terms, the portfolio growth (7.1.5) cannot be ensured in the model world.

In the case of the model following the market but with a few slightly wrong parameter values, the parameter-error view is more appropriate because we now have a way to build a direct link between V^{mkt} and V^{mod}, which is through the Taylor expansion. In this case, the balance of Vanna terms (7.1.4) in (7.1.2) is unnecessary; the balance of Vanna terms has little effect on the error behaviour of the Karasinski price. This is because no matter whether (7.1.4) holds or not, $\left.\frac{\partial^2 V^{mod}}{\partial x \partial S}\right|_{x^{mod}} \delta x \delta S = \sum_k \omega_k \left.\frac{\partial^2 U_k^{mod}}{\partial x \partial S}\right|_{x^{mod}} \delta x \delta S = 0$ will hold as long as $\delta S = 0$. If there is only one parameter x, in the FX market, the balance of Vanna does have the benefit of making three equations in (7.1.2) so the vanilla prices on all three liquid strikes may be exactly produced. In the case of (7.1.2), replacing the balance of Vanna with the balance of ∂Volga$/\partial x$ will make the error of the Karasinski price fourth-order, $V^{mkt} - V^{Kar} = O\left((\delta x)^4\right)$. So (7.1.2) is replaced with

$$\begin{aligned}
\left.\frac{\partial V^{mod}}{\partial x}\right|_{x^{mod}} &= \sum_k \omega_k \left.\frac{\partial U_k^{mod}}{\partial x}\right|_{x^{mod}} \\
\left.\frac{\partial^2 V^{mod}}{\partial x^2}\right|_{x^{mod}} &= \sum_k \omega_k \left.\frac{\partial^2 U_k^{mod}}{\partial x^2}\right|_{x^{mod}} \\
\left.\frac{\partial^3 V^{mod}}{\partial x^3}\right|_{x^{mod}} &= \sum_k \omega_k \left.\frac{\partial^3 U_k^{mod}}{\partial x^3}\right|_{x^{mod}}
\end{aligned} \tag{7.1.6}$$

However, the sensitivity Vanna may be a little easier to compute and the resulting accuracy of using Vanna may be high enough (as will be shown later), so for one parameter x, we can still use (7.1.2).

In the VV method, the model and the market are usually quite different unless the market has a flat Black-Scholes implied vol smile. Although a special case of the broad Karasinski method definition includes the VV method, the Karasinski method is not intended to be used when the model does not follow the market; rather, the Karasinski method is mainly designed to correct calibration errors when the model closely resembles the market with only some slightly wrong model parameter values due to imperfect calibration.

We will demonstrate later that the choice of the parameter-error view for the Karasinski method is proper while the hedging-costs view is not when the model and the market are close with a few parameter values apart. We will show a test on down-out barrier options. If the hedging-costs view is correct, a factor P (e.g. $P_{NoTouch}$) must also be used to adjust

the Karasinski price the same way as the VV method. But if the parameter-error view is correct, no factor P should be needed. It will shown that a factor P should not be added to the Karasinski price formula, confirming our analysis above that the parameter-error view is appropriate for the Karasinski price whereas the hedging-costs view is appropriate for the VV method.

Therefore, in the following discussion, we will follow the parameter-error view for the Karasinski method.

7.1.2. The Karasinski Method in Higher Dimensions. Suppose there are two parameters in a model that are not well calibrated to the market values, and these two parameters are denoted by x_1 and x_2. The parameters x_1 and x_2 move stochastically with time. Assume x_1 and x_2 follow functions of diffusion processes. The Karasinski weights defined in (7.1.2) are good for the case of one parameter; in order to satisfy (7.1.5) for two parameters, we will extend (7.1.2) to

$$\begin{cases} \left.\dfrac{\partial V^{mod}}{\partial x_1}\right|_{x_1^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial U_k^{mod}}{\partial x_1}\right|_{x_1^{mod}}, & \left.\dfrac{\partial V^{mod}}{\partial x_2}\right|_{x_2^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial U_k^{mod}}{\partial x_2}\right|_{x_2^{mod}}, \\ \left.\dfrac{\partial^2 V^{mod}}{\partial x_1^2}\right|_{x_1^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial^2 U_k^{mod}}{\partial x_1^2}\right|_{x_1^{mod}}, & \left.\dfrac{\partial^2 V^{mod}}{\partial x_2^2}\right|_{x_2^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial^2 U_k^{mod}}{\partial x_2^2}\right|_{x_2^{mod}}, \\ \left.\dfrac{\partial^2 V^{mod}}{\partial x_1 \partial x_2}\right|_{x_{1,2}^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial^2 U_k^{mod}}{\partial x_1 \partial x_2}\right|_{x_{1,2}^{mod}}, \end{cases} \tag{7.1.7}$$

In the parameter-error view of the Karasinski method, the balance of Vanna terms is unnecessary, hence they are left out.

In the general case, the Karasinski weights can be computed for N parameters, x_n, $n \leqslant N$, by

$$\begin{cases} \left.\dfrac{\partial V^{mod}}{\partial x_n}\right|_{x_n^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial U_k^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \\ \left.\dfrac{\partial^2 V^{mod}}{\partial x_n^2}\right|_{x_n^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial^2 U_k^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, \\ \left.\dfrac{\partial^2 V^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}} = \displaystyle\sum_k \omega_k \left.\dfrac{\partial^2 U_k^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}, \end{cases} \tag{7.1.8}$$

where $n \leqslant N$, $m \leqslant N$, $n \neq m$.

Suppose M denotes the number of equalities in (7.1.8). For N parameters, the resulting number of equalities M in (7.1.8) can be computed by

$$M = 2N + \frac{N^2 - N}{2} \tag{7.1.9}$$

Note (7.1.8) can also be extended to include higher order sensitivities of V^{mod} w.r.t. x; in some markets where liquid strikes are abundant, this extension will achieve better accuracy. However, we will show that even (7.1.8) using only the Vega terms will be able to achieve excellent results.

7.1.3. Computing ω_k through Pseudo-Inverse when $M \neq L$. Rewrite (7.1.8) for the multi-dimensional Karasinski method in the matrix form:

$$Y = A\omega \tag{7.1.10}$$

where

$$\begin{aligned}
Y &= \left(..., \left.\frac{\partial V^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \left.\frac{\partial^2 V^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, ... \left.\frac{\partial^2 V^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}, ... \right)^T, \\
\omega &= \left(\omega_1, \omega_2, ...\omega_L\right)^T, \\
A &= \begin{pmatrix}
..., \left.\frac{\partial U_1^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \left.\frac{\partial^2 U_1^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, ... \left.\frac{\partial^2 U_1^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}, ... \\
..., \left.\frac{\partial U_2^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \left.\frac{\partial^2 U_2^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, ... \left.\frac{\partial^2 U_2^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}, ... \\
... \\
..., \left.\frac{\partial U_L^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \left.\frac{\partial^2 U_L^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, ... \left.\frac{\partial^2 U_L^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}, ...
\end{pmatrix}^T .
\end{aligned} \tag{7.1.11}$$

When $L \neq M$, to compute ω vector is to find a pseudo-inverse of A, which is a $M \times L$ matrix.

If $L > M$, the system of equations are under-determined, there are many solutions to (7.1.10). To find one solution, the square of the Euclidean norm of ω, $||\omega||^2$, is minimised. This the least-norm pseudo-inverse [47, 10, 2]. This is the same as what Karasinski has suggested.

If $M > L$, the system of equations are over-determined, there are no exact solutions to (7.1.10). So a solution can only be found in a least-squares sense by minimising the square of the Euclidean norm of $||A\omega - Y||^2$. This is the Moore-Penrose pseudo-inverse [74, 81, 2].

The formula ω can be computed by the following formula (see [47, 10]):

$$\omega = \begin{cases} A^T \left(AA^T\right)^{-1} B, & M < L, \\ \left(A^T A\right)^{-1} A^T B, & M > L. \end{cases} \tag{7.1.12}$$

Including the case of $M = L$, the formula for ω can be written as

$$\omega = \begin{cases} A^T \left(AA^T\right)^{-1} B, & M < L, \\ (A)^{-1} B & M = L \\ \left(A^T A\right)^{-1} A^T B, & M > L. \end{cases} \tag{7.1.13}$$

7.1.4. Model Consistency of the Karasinski Method with the Liquid Prices U_k^{mkt}. A fundamental requirement of a model is the ability to fit the liquid market vanilla prices U_k^{mkt}. We will consider whether the general case, the Karasinski method has the required price consistency on the liquid vanillas.

When the number of vanillas and the number of equalities in (7.1.8) are equal, $L = M = 2N + \dfrac{N^2 - N}{2}$, the consistency can be proven. Following (5.2.23), the sensitivities in the equalities in (7.1.8) are written as D_1, D_2,..., D_M. Without loss of generality, suppose that we try to price the vanilla U_1 with the Karasinski method. Hence (7.1.2) becomes

$$\begin{aligned} D_1 U_1^{mod} &= \omega_1 D_1 U_1^{mod} + \omega_2 D_1 U_2^{mod} + ... + \omega_L D_1 U_L^{mod} \\ D_2 U_1^{mod} &= \omega_1 D_2 U_1^{mod} + \omega_2 D_2 U_2^{mod} + ... + \omega_L D_2 U_L^{mod} \\ &... \\ D_M U_1^{mod} &= \omega_1 D_M U_1^{mod} + \omega_2 D_M U_2^{mod} + ... + \omega_L D_M U_L^{mod} \end{aligned} \tag{7.1.14}$$

to which $\omega_1 = 1$, $\omega_2 = 0$,... $\omega_L = 0$ is obviously a solution. The sensitivities D_1, D_2,..., D_M are evaluated at the current model parameters.

Assume further the matrix

$$\left(\begin{array}{cccc} D_1 U_1^{mod} & D_1 U_2^{mod} & ... & D_1 U_L^{mod} \\ D_2 U_1^{mod} & D_2 U_2^{mod} & ... & D_2 U_L^{mod} \\ ... & & ... & \\ D_M U_1^{mod} & D_M U_2^{mod} & ... & D_M U_L^{mod} \end{array} \right) \Bigg|_{L=M} \tag{7.1.15}$$

is invertible.

Therefore the weights ω_k admit a unique solution. Under this assumption, $\omega_1 = 1$, $\omega_2 = 0$,... $\omega_L = 0$ is the unique solution. Therefore, the Karasinski price for U_1 is $U_1^{mod} + 1\left(U_1^{mkt} - U_1^{mod}\right) = U_1^{mkt}$, a perfect fit of the market price of the liquid vanilla. For the other vanillas U_2,...,U_L, the same proof applies.

In the case of $L \neq M$, depending on the choice of the pseudo-inverse method for the matrix A in (7.1.10), the solution $\omega_{k'} = 1$, $\omega_{k \neq k'} = 0$ for a liquid vanilla $U_{k'}$ is not

necessarily guaranteed. In the case of $L < M$, however, the Moore-Penrose pseudo-inverse [74, 81] ensures the solution $\omega_{k'} = 1$, $\omega_{k \neq k'} = 0$ for a liquid vanilla $U_{k'}$ hold as well because by the definition of the Moore-Penrose pseudo-inverse, the solution $\omega_{k'} = 1$, $\omega_{k \neq k'} = 0$ for a liquid vanilla $U_{k'}$ will ensure $||A\omega - Y||^2 = 0$ on the strike of $U_{k'}$. In the case of $L > M$, the solution of $\omega_{k'} = 1$, $\omega_{k \neq k'} = 0$ for a liquid vanilla $U_{k'}$ is usually not guaranteed for all liquid vanillas; we will show later that although the error of the Karasinski price increases a little as a result of the Karasinski method not producing all of the liquid vanilla prices, the application of the pseudo-inverse method has very limited negative influence on the accuracy of the Karasinski method.

7.1.5. Mapping the Black-Scholes Implied Vol Smile to a Flat x Smile. Suppose the model closely follows the market; there is only one model parameter x that has an incorrect value due to imperfect calibration and other than this parameter all the other parameters of the chosen model are well calibrated. Following the same error analysis in the VV chapter, for $M \leqslant L$, we obtain for the Karasinski method

$$V^{mkt} - V^{Kar} \approx O\left((\delta x)^3\right) \tag{7.1.16}$$

where δx is the difference between the market and model x values. If $M > L$, the equalities in (7.1.8) will not be exact, hence the breaking of the cubic error behaviour.

We have shown earlier that the VV method will see the cubic error convergence $V^{mkt} - V^{VV} \approx O\left((\delta\sigma)^3\right)$ when the market Black-Scholes vol smile is flat and the Black-Scholes model is a close reflection of the market with a small difference in its vol. If x is viewed as a model implied "x smile" in the same vein as the Black-Scholes implied volatility, the x smile is flat even when the Black-Scholes implied vol smile may be not. This is an advantage of the Karasinski method in the case of the model with a smile closely following the market but having a wrong x value: while the VV method needs a flat market Black-Scholes implied vol smile to see cubic error, the Karasinski may have the cubic error even when the market Black-Scholes implied vol smile is not flat.

Similarly, assuming that the model follows the market but with slightly wrong x_1, x_2, ..., x_n,..., x_N parameter values, for $M \leqslant L$, we will see the error

$$V^{mkt} - V^{Kar} \approx \sum O\left(\delta x_n \delta x_m \delta x_l\right) \tag{7.1.17}$$

where δx_n, δx_m, δx_l, $1 \leqslant n$, m, $l \leqslant N$, is the difference between the market and model x_n, x_m, x_l values.

7.2. The Vega-Karasinski Method for $M > L$

If $M > L$, we rely on the pseudo-inverse to find the weights ω_k, which may lower the accuracy of the Karasinski method. As a result of the pseudo-inverse, the error behaviours described by (7.1.16) and (7.1.17) are not cubic; instead, they will be first-order, $V^{mkt} - V^{Kar} \approx O(\delta x)$ and $V^{mkt} - V^{Kar} \approx \sum_{n=1}^{N} O(\delta x_n)$, respectively.

In the FX market, if we correct the model price error due to three incorrect model parameters x_1, x_2 and x_3 with the first order sensitivities only, we will have

$$\begin{cases} \dfrac{\partial V^{mod}}{\partial x_1} &= \displaystyle\sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_1} \\ \dfrac{\partial V^{mod}}{\partial x_2} &= \displaystyle\sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_2} \\ \dfrac{\partial V^{mod}}{\partial x_3} &= \displaystyle\sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_3} \end{cases} \tag{7.2.1}$$

which gives the quadratic error

$$V^{mkt} - V^{Kar} \approx \sum_{n=1}^{3} O\left((\delta x_n)^2\right) + \sum_{n \neq m} O(\delta x_n \delta x_m) \tag{7.2.2}$$

while still allowing the Karasinski formula to fit exactly the market prices of all three liquid strikes. For easy reference, we will call this the Vega-Karasinski method because it uses only the first-order sensitivities.

When $L = N$ in a general case, the weights for the Vega-Karasinski method are computed by

$$\frac{\partial V^{mod}}{\partial x_n} = \sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_n}$$

where $1 \leqslant n \leqslant N$. Similarly, the error will follow $V^{mkt} - V^{Kar} \approx \sum_{n=1}^{N} O\left((\delta x_n)^2\right) + \sum_{n \neq m} O(\delta x_n \delta x_m)$.

Although the Karasinski method with a pseudo-inverse balances the first-order and second-order sensitivities, both the first and the second sensitivities are balanced in an approximate sense. On the other hand, when $M > L$, the Vega-Karasinski method leaves out the second-order sensitivities but makes a better balance of the first-order sensitivities. Comparing the Vega-Karasinski method with the Karasinski method with a pseudo-inverse,

we find that the error behaviour of the Vega-Karasinski method appears to be better; the error of Vega-Karasinski method is second-order compared the first-order Karasinski method with a pseudo-inverse when $M > L$. Later we will show in numerical examples that the Vega-Karasinski method is better in accuracy than the Karasinski method with a pseudo inverse when $M > L$.

In the FX market ($L = 3$), if there are only two incorrect model parameters x_1 and x_2, the Karasinski weights ω_k computed by the equations

$$\begin{cases} \dfrac{\partial V^{mod}}{\partial x_1} &= \displaystyle\sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_1} \\ \dfrac{\partial V^{mod}}{\partial x_2} &= \displaystyle\sum_k \omega_k \frac{\partial U_k^{mod}}{\partial x_2} \\ \dfrac{\partial^2 V^{mod}}{\partial x_1^2} &= \displaystyle\sum_k \omega_k \frac{\partial^2 U_k^{mod}}{\partial x_1^2} \end{cases} \tag{7.2.3}$$

will ensure that the quadratic error (7.2.2) continues to hold. We will call the Karasinski method based on the weights computed by (7.2.3) the Vega-Karasinski method as well, because the first-order sensitivities with respect to incorrect model parameters are balanced and the second-order sensitivities are only partially balanced.

7.3. Treat a Greek as the Target Instrument

Currently there are no published or known methods to compute Greeks with the VV method as Carr and Wystup [104] indicated. In practice, the Black-Scholes Greeks are usually used while the VV method is used to produce prices. It is problematic because the Greeks are often priced on the wrong vols as the usual candidate for the model vol is the at-the-money vol and the Black-Scholes dynamics might not be a close reflection of the market. If we directly take derivatives of the VV price, we will encounter the challenge of computing the Greeks of market vanilla prices. The sensitivity of the VV price of an option V with respect to a model parameter y satisfies

$$\begin{aligned} \frac{\partial V^{VV}}{\partial y} &= \left.\frac{\partial V^{BS}}{\partial y}\right|_{\sigma_{ATM}} + \sum_k \frac{\partial \omega_k}{\partial y}\left(U_k^{mkt} - U_k^{BS}|_{\sigma_{ATM}}\right) \\ &\quad + \sum_k \omega_k \left(\frac{\partial U_k^{mkt}}{\partial y} - \left.\frac{\partial U_k^{BS}}{\partial y}\right|_{\sigma_{ATM}}\right) \end{aligned} \tag{7.3.1}$$

While the model Greeks can be computed, we have no idea what the market Greeks of the vanillas are. The lack of a method to compute Greeks using the concepts of the VV

and Karasinski approaches mainly involves the inability to obtain derivatives of market prices.

7.3.1. Price a Greek like a Target Instrument in the Karasinski Method. We find that this issue can be circumvented if we treat the Greek to be priced as the target instrument in the Karasinski formula and change the definition of the hedging weights in the Karasinski methods. Treating a Greek as the target instrument means that we regard the Greek as an option price in the Karasinski Method. Essentially, by this approach, the Greek of an option price V takes the position of V in the Karasinski formula. The meaning of this approach will be made clear with the formula and explanation laid out below.

First example, the Karasinski formula for computing a first-order sensitivity $\dfrac{\partial V^{Kar}}{\partial y}$ is

$$\frac{\partial V^{Kar}}{\partial y} = \left.\frac{\partial V^{mod}}{\partial y}\right|_{x^{mod}} + \sum_k \omega_k^y \left(U_k^{mkt} - \left.U_k^{mod}\right|_{x^{mod}}\right) \tag{7.3.2}$$

where the weights ω_k^y are computed from

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial V^{mod}}{\partial y}\right)\Big|_{x^{mod}} &= \sum_k \omega_k^y \left.\frac{\partial U_k^{mod}}{\partial x}\right|_{x^{mod}} \\
\frac{\partial^2}{\partial x^2}\left(\frac{\partial V^{mod}}{\partial y}\right)\Big|_{x^{mod}} &= \sum_k \omega_k^y \left.\frac{\partial^2 U_k^{mod}}{\partial x^2}\right|_{x^{mod}} \\
\frac{\partial^3}{\partial x^3}\left(\frac{\partial V^{mod}}{\partial y}\right)\Big|_{x^{mod}} &= \sum_k \omega_k^y \left.\frac{\partial^3 U_k^{mod}}{\partial x^3}\right|_{x^{mod}}
\end{aligned} \tag{7.3.3}$$

The VV formula for computing a first-order sensitivity $\dfrac{\partial V^{VV}}{\partial y}$ is

$$\frac{\partial V^{VV}}{\partial y} = \left.\frac{\partial V^{BS}}{\partial y}\right|_{\sigma_{ATM}} + \sum_k \omega_k^y \left(U_k^{mkt} - \left.U_k^{BS}\right|_{\sigma_{ATM}}\right)$$

where the weights ω_k^y are computed from

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial V^{BS}}{\partial y}\right)\Big|_{\sigma_{ATM}} &= \sum_k \omega_k^y \left.\frac{\partial U_k^{BS}}{\partial x}\right|_{\sigma_{ATM}} \\
\frac{\partial^2}{\partial x^2}\left(\frac{\partial V^{BS}}{\partial y}\right)\Big|_{\sigma_{ATM}} &= \sum_k \omega_k^y \left.\frac{\partial^2 U_k^{BS}}{\partial x^2}\right|_{\sigma_{ATM}} \\
\frac{\partial^2}{\partial x \partial S}\left(\frac{\partial V^{BS}}{\partial y}\right)\Big|_{\sigma_{ATM}} &= \sum_k \omega_k^y \left.\frac{\partial^2 U_k^{BS}}{\partial x \partial S}\right|_{\sigma_{ATM}}
\end{aligned}$$

Similarly, the Karasinski formula for a second-order sensitivity $\left.\dfrac{\partial^2 V^{mod}}{\partial y \partial z}\right|_{x^{mod}}$ is

$$\frac{\partial^2 V^{Kar}}{\partial y \partial z} = \left.\frac{\partial^2 V^{mod}}{\partial y \partial z}\right|_{x^{mod}} + \sum_k \omega_k^{yz} \left(U_k^{mkt} - \left. U_k^{mod} \right|_{x^{mod}} \right) \tag{7.3.4}$$

where the weights ω_k^{yz} are computed from

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\left.\frac{\partial^2 V^{mod}}{\partial y \partial z}\right)\right|_{x^{mod}} &= \sum_k \omega_k^{yz} \left.\frac{\partial U_k^{mod}}{\partial x}\right|_{x^{mod}} \\
\frac{\partial^2}{\partial x^2}\left(\left.\frac{\partial^2 V^{mod}}{\partial y \partial z}\right)\right|_{x^{mod}} &= \sum_k \omega_k^{yz} \left.\frac{\partial^2 U_k^{mod}}{\partial x^2}\right|_{x^{mod}} \\
\frac{\partial^3}{\partial x^3}\left(\left.\frac{\partial^2 V^{mod}}{\partial y \partial z}\right)\right|_{x^{mod}} &= \sum_k \omega_k^{yz} \left.\frac{\partial^3 U_k^{mod}}{\partial x^3}\right|_{x^{mod}}
\end{aligned} \tag{7.3.5}$$

Our approach of computing Greeks with the Karasinski model is a very straightforward idea. If the model closely follows the market and x is the only model parameter that has a wrong value, the error of a Karasinski sensitivity is cubic.

Suppose G denotes a Greek to be priced. Similarly, for the Karasinski method, the formula for pricing the Greek G for $N > 1$ is

$$G^{Kar} = \left. G^{mod} \right|_{x^{mod}_{1,\dots N}} + \sum_k \omega_k \left(U_k^{mkt} - \left. U_k^{mod} \right|_{x^{mod}_{1,\dots N}} \right)$$

and the weights ω_k satisfy

$$\begin{cases}
\left.\dfrac{\partial G^{mod}}{\partial x_n}\right|_{x_n^{mod}} = \displaystyle\sum_k \omega_k \left.\frac{\partial U_k^{mod}}{\partial x_n}\right|_{x_n^{mod}}, \\
\left.\dfrac{\partial^2 G^{mod}}{\partial x_n^2}\right|_{x_n^{mod}} = \displaystyle\sum_k \omega_k \left.\frac{\partial^2 U_k^{mod}}{\partial x_n^2}\right|_{x_n^{mod}}, \\
\left.\dfrac{\partial^2 G^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}} = \displaystyle\sum_k \omega_k \left.\frac{\partial^2 U_k^{mod}}{\partial x_n \partial x_m}\right|_{x_{n,m}^{mod}}.
\end{cases}$$

where $n \leqslant N$, $m \leqslant N$, $n \neq m$.

If the model is the Black-Scholes model, $N = 1$ and $L = 3$, we will also call G^{Kar} the VV Greek G^{VV}.

7.4. Fast Linear Model Calibration Coupled with the Karasinski Method.

When a complex model is frequently but imperfectly calibrated, we regard this model as a model closely following the market but with slightly wrong model parameter estimates. In this case, the Karasinski method can work with a fast linear calibration scheme to offer the best performance and speed.

Assuming the vanilla prices U^{mkt} and U^{mod} are close with the parameters x_n, $1 \leqslant n \leqslant N$, having slightly different values between the market and the model, we take Taylor's expansion of $U^{mkt} - U^{mod}$ on the liquid strikes and obtain

$$U_k^{mkt} - U_k^{mod} \approx \sum_{n=1}^{N} \left(\frac{\partial U_k^{mod}}{\partial x_n} \delta x_n + \frac{1}{2} \frac{\partial^2 U_k^{mod}}{\partial x_n^2} (\delta x_n)^2 \right) + \sum_{n \neq m} \frac{\partial^2 U_k^{mod}}{\partial x_n \partial x_m} \delta x_n \delta x_m \qquad (7.4.1)$$

where δx_n denotes the x_n parameter difference between the market and the model. We know $U^{mkt} - U^{mod}$ and ω_k for each strike, so the question of model calibration is to compute δx_n through a system of quadratic equations. Solving the quadratic equations may be easier than the traditional calibration method.

If the market-model parameter differences δx_n are small enough, (7.4.1) can be further reduced to the first-order linear form:

$$U_k^{mkt} - U_k^{mod} \approx \sum_{n=1}^{N} \left(\frac{\partial U_k^{mod}}{\partial x_n} \delta x_n \right) \qquad (7.4.2)$$

which is a system of linear equations that can be easily solved.

We can use the solutions from (7.4.2) to estimate the solutions of (7.4.1) approximately by solving another system of linear equations. Assume that $\delta x_n^{(1)}$ are the solutions obtained from the linear equations (7.4.2) and that $\delta x_n^{(2)}$ are the solutions obtained from the quadratic equations (7.4.1). Assume further that $\delta x_n^{(2)}$ can be expressed by

$$\delta x_n^{(2)} = \delta x_n^{(1)} + \delta x_n^{(0)} \qquad (7.4.3)$$

Substituting (7.4.3) into (7.4.1) gives

$$\begin{aligned} U_k^{mkt} - U_k^{mod} \quad \approx \quad & \sum_{n=1}^{N} \left(\frac{\partial U_k^{mod}}{\partial x_n} \left(\delta x_n^{(1)} + \delta x_n^{(0)} \right) + \frac{1}{2} \frac{\partial^2 U_k^{mod}}{\partial x_n^2} \left(\delta x_n^{(1)} \right)^2 \right) \\ & + \sum_{n \neq m} \frac{\partial^2 U_k^{mod}}{\partial x_n \partial x_m} \delta x_n^{(1)} \delta x_m^{(1)} \end{aligned} \qquad (7.4.4)$$

where the quadratic terms containing $\delta x_n^{(0)}$ in (7.4.4) have been ignored for they are far smaller than the first-order terms $\dfrac{\partial U_k^{mod}}{\partial x_n} \delta x_n^{(0)}$. After $\delta x_n^{(1)}$ have been solved through the system of linear equations (7.4.2), to obtain $\delta x_n^{(0)}$ from (7.4.4) is a simple task of solving another system of linear equations . Therefore, to solve the quadratic equations (7.4.1) approximately is to solve two systems of linear equations, which is very easy and fast. Once $\delta x_n^{(0)}$ are computed, $\delta x_n^{(2)}$ can be computed with (7.4.3). This method can potentially be

extended to even higher orders, but we will show later that the second-order calibration is already very accurate for the Karasinski method and even when used on its own.

For easy reference, we will call the first-order calibration the FLF (fast linear first-order) calibration, and the second-order calibration the FLQ (fast linear quadratic) calibration.

Both (7.4.1) and (7.4.2) have the high-order terms cut off, hence not perfect. However, the Karasinski method does not need extremely high accuracy in model parameters; it just needs a model that is reasonably close to the market. Later we will see that 10%, 20% relative errors in model parameters still make the Karasinski prices highly accurate in our test examples.

If the question is to maintain a calibration that is not far behind the market and is good to use with the Karasinski method, we can either solve the equations (7.4.2) to conduct a first-order calibration (FLF), or use the combination of (7.4.2), (7.4.4) and (7.4.3) to give a second-order calibration (FLQ), both of which can be quickly computed through linear equations. Every time the Karasinski method is run, the first-order or the quadratic calibration can be used to bring the x_n roughly to the market values. The resulting model price error because of the imperfection in x_n can be corrected by the Karasinski method.

Therefore, the Karasinski method and the FLF or FLQ calibration schemes can act in partnership to offer great efficiency and speed.

7.5. Performance Analysis

7.5.1. Inverse of Square Matrix A, $(M = L = 3)$ - CEV Market and Model Differ by One Parameter $(N = 1)$. As we have analysed, to compute the Karasinski weights ω_k, the matrix A in (7.1.13) needs to be either inverted or pseudo-inverted depending on M and L.

When $M = L$, the matrix A in (7.1.13) can be simply inverted. A test based on the CEV model described by (5.5.3) is performed. Except the model CEV parameter σ^{mod} for the Karasinski method that is specified in the plots of Figure (7.5.1), the parameters for the CEV test are the same as those used for Figure (5.5.11); the model CEV parameter σ^{mod} is assumed to have different values from the market CEV parameter σ^{mkt} due to the imperfection in calibration in these plots. The top plot of Figure (7.5.1) shows the relative error between the CEV market price and the Karasinski prices for different strikes under various parameter estimation errors in σ^{mod}; the bottom plot of Figure (7.5.1) shows

the relative error between the CEV market price and the CEV model, the VV, and the Karasinski prices for a European call with the strike $K = 0.5\,(K_2 + K_3)$. In the top plot of Figure (7.5.1), the European call prices are compared for $K > S$ and the European put prices are compared for $K < S$. The top plot of Figure (7.5.1) reveals that the Karasinski prices with the relative σ^{mod} estimation errors less than 10% demonstrate much smaller errors than those of the VV prices, in both the interpolation and extrapolation ranges of the strike. The bottom plot of Figure (7.5.1) shows cubic convergence (with Vanna terms balanced) and fourth-order convergence (with ∂Volga/∂x terms balanced) of the Karasinski price; it can be seen that the Karasinski price is a more accurate price than the VV price if the relative σ^{mod} estimation error is smaller than a certain level, such as 10% or 20% in this plot. Figure (7.5.1) shows that replacing the balance of Vanna terms with the balance of ∂Volga/∂x achieves higher accuracy, which implies that the balance of Vanna terms is not essential.

7.5.2. Pseudo-Inverse of Matrix A.

7.5.2.1. *Over-Determined* (7.1.10) - $M = 5 > L = 3$. The top plot of Figure (7.5.2) shows the relative error between the SABR market price and the Karasinski prices for different strikes under various parameter estimation errors in the SABR model parameters ρ^{mod} and ν^{mod}; the bottom plot of Figure (7.5.2) shows the relative error between the SABR market price and the Karasinski prices for a European call with the strike $K = 0.5\,(K_2 + K_3)$. In the top plot of Figure (7.5.2), the European call prices are compared for $K > S$ and the European put prices are compared for $K < S$. Except that the SABR model parameters ρ^{mod} and ν^{mod} are specified in the plots, the parameters for Figure (7.5.2) are the same as those of Figure (5.5.14). Since there are two incorrect model parameters, $N = 2$, we have $M = 5 > L = 3$, hence having a system of over-determined equations in (7.1.10). With the system of over-determined equations, as shown on Figure (7.5.2), the pseudo-inverse of the matrix A makes the Karasinski method perfectly match the market vanilla prices on the liquid strikes.

Figure (7.5.2) shows that the pseudo-inverse of the matrix A works very well; with the model parameter estimation errors within 10%, the resulting relative error between the SABR market price and the Karasinski price is in the order around 10^{-4} to 10^{-3}. The bottom plot of Figure (7.5.2) shows that even with 20% model parameter estimation errors, the relative error of the Karasinski price is roughly -1.2×10^{-3}, which is still very small.

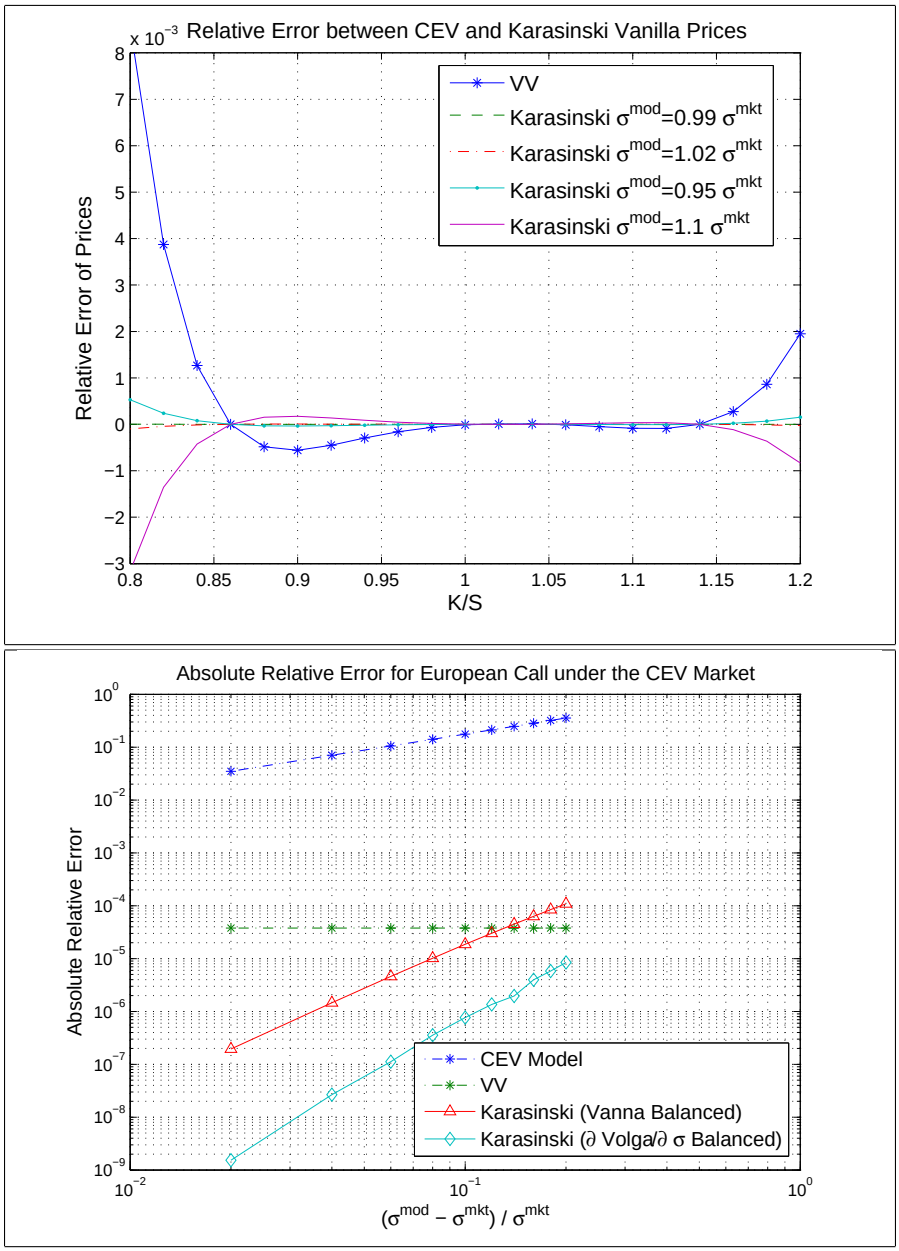

FIGURE 7.5.1. Cubic Convergence (with Vanna Terms Balanced) and Fourth-order Convergence (with ∂Volga/∂x Terms Balanced) of the Karasinski Price , $M = L = 3$

The bottom plot of Figure (7.5.2) also shows that with the pseudo-inverse on the over-determined equations, the error behaviour of the Karasinski price will no longer be cubic; the error behaviour instead demonstrates a first-order convergence with respect to changes in the SABR model parameters ρ^{mod} and ν^{mod}.

7.5.2.2. *Further Over-Determined* (7.1.10) - $M = 9 > L = 3$ *and* $M = 20 > L = 3$. Figure (7.5.3) shows the relative error between the Heston market price and the Karasinski prices for different strikes under various parameter estimation errors in the Heston model

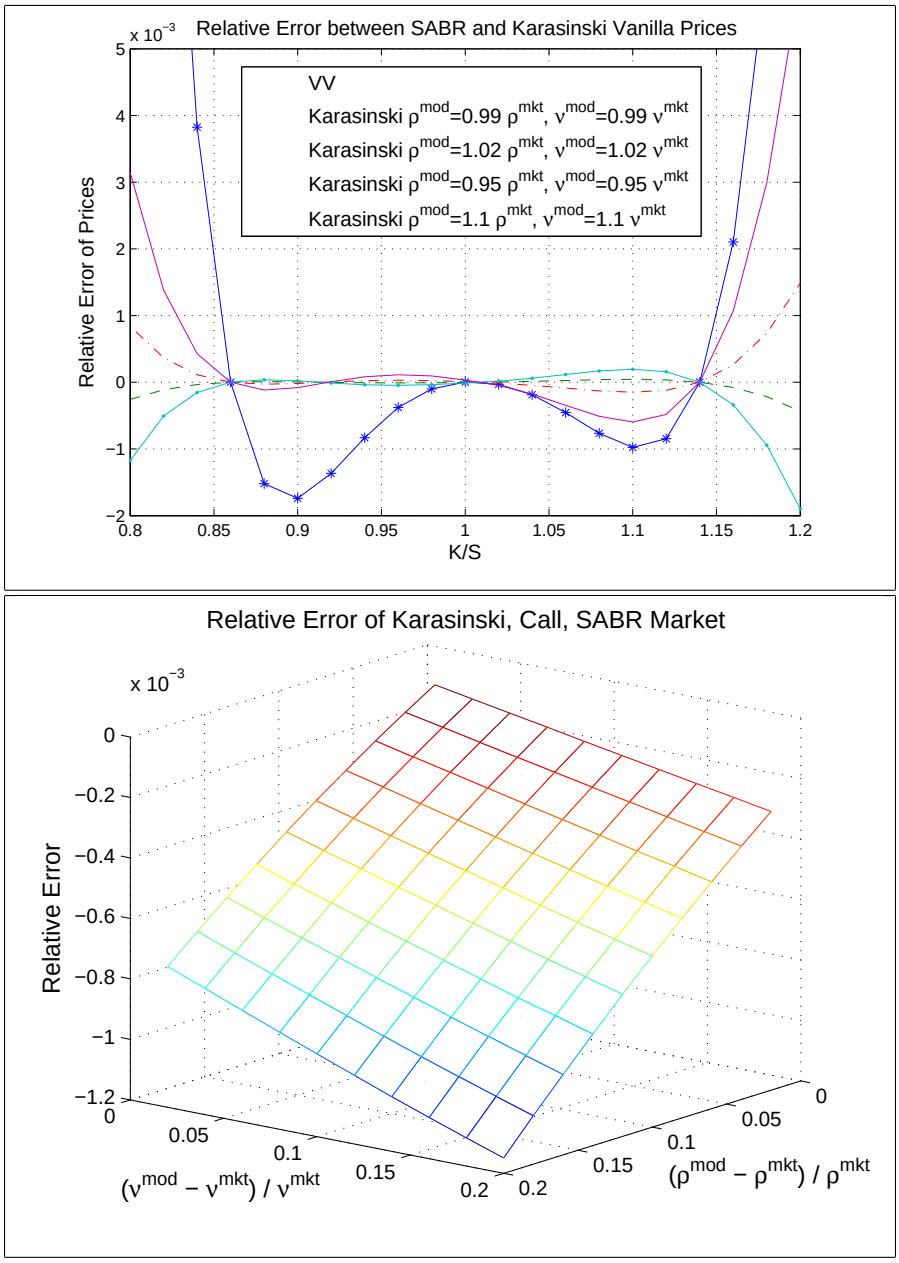

FIGURE 7.5.2. SABR Market and Model Differ by Two Parameters ($N = 2$)

parameters θ^{mod}, ξ^{mod}, and ρ^{mod}. In Figure (7.5.3), the European call prices are compared for $K > S$ and the European put prices are compared for $K < S$. Except that the Heston model parameters θ^{mod}, ξ^{mod}, and ρ^{mod} are specified in the plots, the parameters for Figure (7.5.3) are the same as those of Figure (5.5.3). Because of $N = 3$, we have $M = 9 > L = 3$, hence having a more over-determined system of equations (7.1.10) than the SABR case described in the above section, which has $M = 5$ and $L = 3$. A comparison between Figure (7.5.3) and the top plot of Figure (7.5.2) shows that as the system of equations (7.1.10) becomes more over-determined, the relative error of the Karasinski price will increase; the

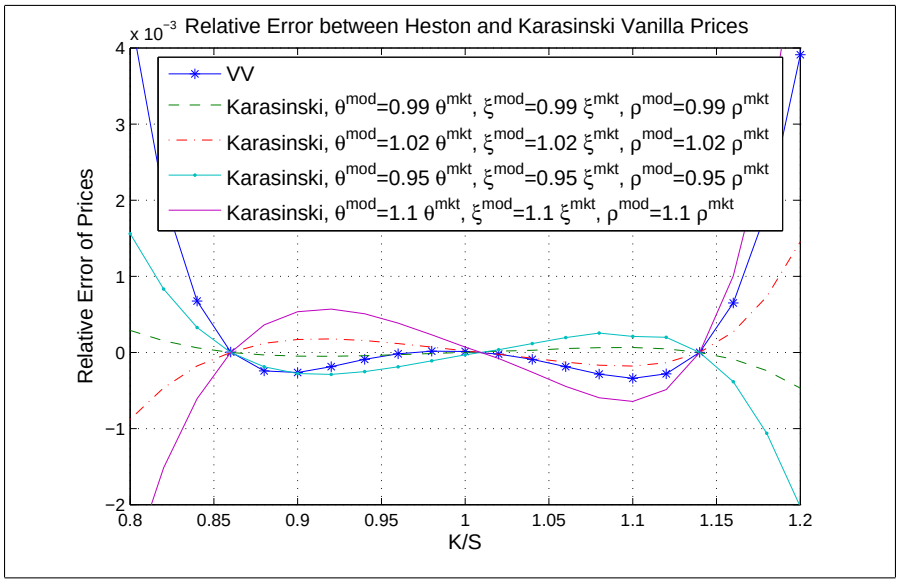

FIGURE 7.5.3. Heston Market and Model Differ by Three Parameters ($N = 3$)

relative error curve with a 10% parameter estimation error in the top plot of Figure (7.5.2) is flat and close to zero between the liquid strikes but the relative error curve with a 10% parameter estimation error in Figure (7.5.3) shows slightly larger errors. However, Figure (7.5.3) shows that the pseudo-inverse still works quite well with the resulting relative errors of the Karasinski prices in the order around 10^{-4} and 10^{-3} when the relative parameter estimation errors are within 10%.

Inspired by the success of the Moore-Penrose pseudo-inverse demonstrated by Figures (7.5.2) and (7.5.3), we will extend the above Heston result one step further by assuming that all five parameters of the Heston model, κ^{mod}, θ^{mod}, ξ^{mod}, ρ^{mod} and ν_0^{mod} are slightly wrong compared to the Heston market parameter values. Except the model parameters κ^{mod}, θ^{mod}, ξ^{mod}, ρ^{mod} and ν_0^{mod} specified in Figure (7.5.4), all the other parameters are the same as those for Figure (7.5.3). Because of $N = 5$, we have $M = 20 > L = 3$. Figure (7.5.4) shows very little changes from Figure (7.5.3).

This shows the the pseudo-inverse method is highly robust with the Karasinski method. With only three liquid vanillas, the Karasinski method can reliably correct its model price computed with at least five wrong model parameters.

7.5.2.3. *Under-Determined* (7.1.10) - $M = 3 < L = 5$. The parameters for Figure (7.5.5) are the same as those for the top plot of Figure (7.5.1) except that the two error curves marked with $L = 5$ in Figure (7.5.5) assume two additional liquid strikes $K_0 = K_1 - 0.3$ and $K_4 = K_3 + 0.3$ are also available in this CEV market, hence $L = 5$. We have

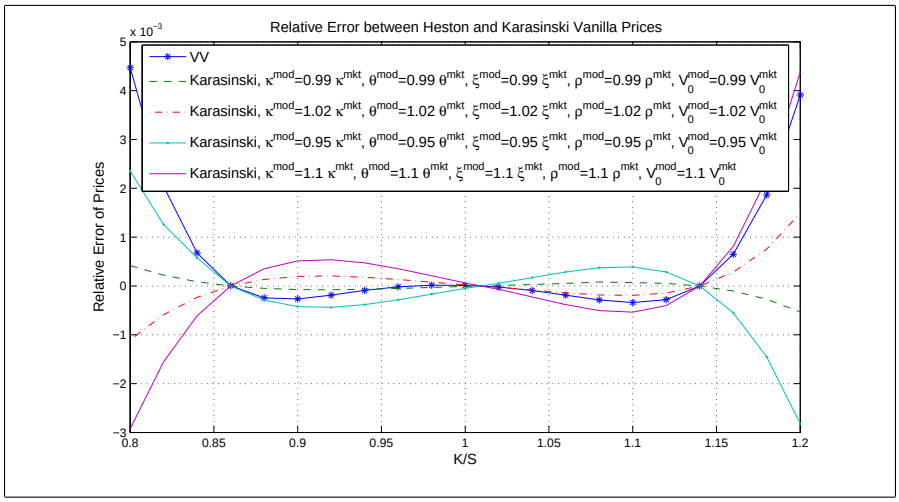

FIGURE 7.5.4. Heston Market and Model Differ by Five Parameters ($N = 5$)

known that the market vanilla prices on the liquid strikes are produced by the Karasinski method for $M \geqslant L$, provided the Moore-Penrose pseudo-inverse [74, 81] is used for $M > L$. However, when $M < L$, as shown in Figure (7.5.5), the market vanilla prices on the liquid strikes are not necessarily produced by the Karasinski method.

Figure (7.5.5) shows that the relative errors of the Karasinski method with a system of under-Determined equations ($M = 3$, $L = 5$) in (7.1.10) are larger than those of the Karasinski method with a square matrix A ($L = M = 3$) in (7.1.10). However, the error difference between the case of $M = 3$, $L = 5$ and the case of $L = M = 3$ is very small; the errors for both cases are roughly in the same order.

This suggests that the least-norm pseudo-inverse used for under-determined (7.1.10) has a very limited impact on the accuracy of the Karasinski method.

The other side of the same issue is that we can simply use just enough number of liquid vanillas to make $L = M$, even if more liquid vanillas are available. For a strike K, we can use the liquid strikes near K to make $L = M$ in the Karasinski method. With increased computational costs and no improvement on error, to use more liquid vanillas than equations ($L > M$) appears to offer no clear benefit.

7.5.3. Performance of the Vega-Karasinski Method. The parameters for Figure (7.5.6) are the same as those for Figure (7.5.3). Figure (7.5.6) shows the relative error of the Vega-Karasinski price whereas Figure (7.5.3) shows the relative error of the Karasinski price through pseudo-inverse. A comparison between Figure (7.5.6) and Figure (7.5.3) clearly shows the accuracy advantage of the Vega-Karasinski method. In Figure (7.5.6),

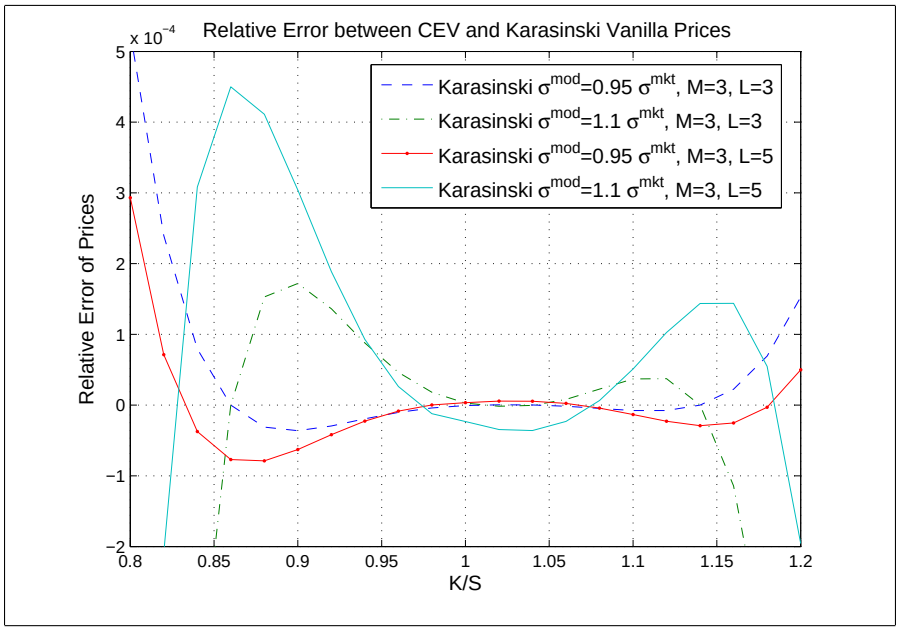

FIGURE 7.5.5. Comparison of Relative Vanilla Price Error between the Case of $M = L = 3$ and the Case of $M = 3$, $L = 5$ under a CEV Market

the relative errors of the Vega-Karasinski method stay flat and close to zero between liquid strikes.

To see the accuracy of the Vega-Karasinski method more clearly, we compare the relative errors of the Vega-Karasinski price and the Karasinski price for a European call with the strike $0.5\,(K_2 + K_3)$ with respect to the Heston model parameter changes in Figure (7.5.7). Except the Heston model parameter changes as shown in the plots, all other parameters of Figure (7.5.7) are the same as those used for Figure (7.5.6). The left plots of Figure (7.5.7) show the relative error of the Karasinski price and the right plots of Figure (7.5.7) show the relative error of the Vega-Karasinski price. The top plots of Figure (7.5.7) show the relative error with respect to model ξ, θ changes and the bottom plots of Figure (7.5.7) show the relative error with respect to model ξ, ρ changes. Figure (7.5.7) shows that the Vega-Karasinski price is roughly 10 to 20 times more accurate than the Karasinski price obtained through pseudo-inverse; the Vega-Karasinski price is much more accurate because the first-order error has been removed in the Vega-Karasinski price. Due to the existence of the first-order error component in the Karasinski price obtained through pseudo-inverse, the left plots of Figure (7.5.7) show linear error behaviours; because the error of the Vega-Karasinski price is a sum of quadratic terms, the right plots of Figure (7.5.7) have curvy surfaces.

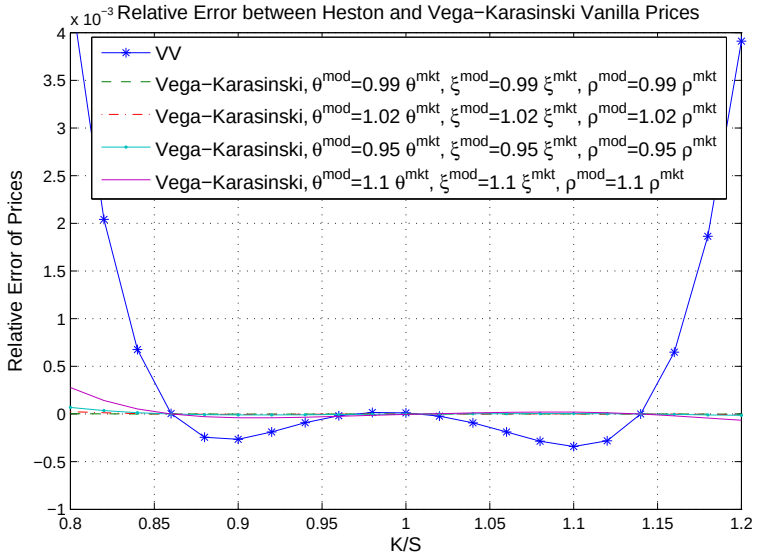

FIGURE 7.5.6. Heston Market and Model Differ by Three Parameters ($N = 3$)

Figure (7.5.8) shows the performance of the Vega-Karasinski method described by (7.2.3) ($L = 3$ and $N = 2$) under a SABR market. The parameters for Figure (7.5.8) are the same as those for the bottom plot of Figure (7.5.2). A comparison between Figure (7.5.8) and the bottom plot of Figure (7.5.2) shows that the Vega-Karasinski method demonstrated by (7.2.3) is more than 10 times more accurate than the Karasinski method demonstrated in the bottom plot of Figure (7.5.2).

Figures (7.5.6) and (7.5.7) demonstrate the importance of fully cancelling out the lower-order error component before taking the higher-order error components into account.

7.5.4. The Karasinski Greeks. Figure (7.5.9) shows the relative error of the Black-Scholes, SABR, VV and Vega-Karasinski prices of a European call option with the strike $0.5\,(K_2 + K_3)$ under a SABR model market. The parameters of Figure (7.5.9) are the same as Figure (7.5.2). The Black-Scholes Greek prices are computed at σ_{ATM}; the SABR model parameters ρ and ν are different from the market values by 10%; the VV prices are computed by treating the Black-Scholes Delta and Gamma as the target instruments for the VV method; the Vega-Karasinski prices are computed by treating the SABR model Delta and Gamma as the target instruments for the Vega-Karasinski method.

Figure (7.5.9) shows that the relative error of the Vega-Karasinski Greek values are very close to zero and far better than the other displayed choices of values. The VV Greek values are not as accurate as the Vega-Karasinski values because the Black-Scholes model

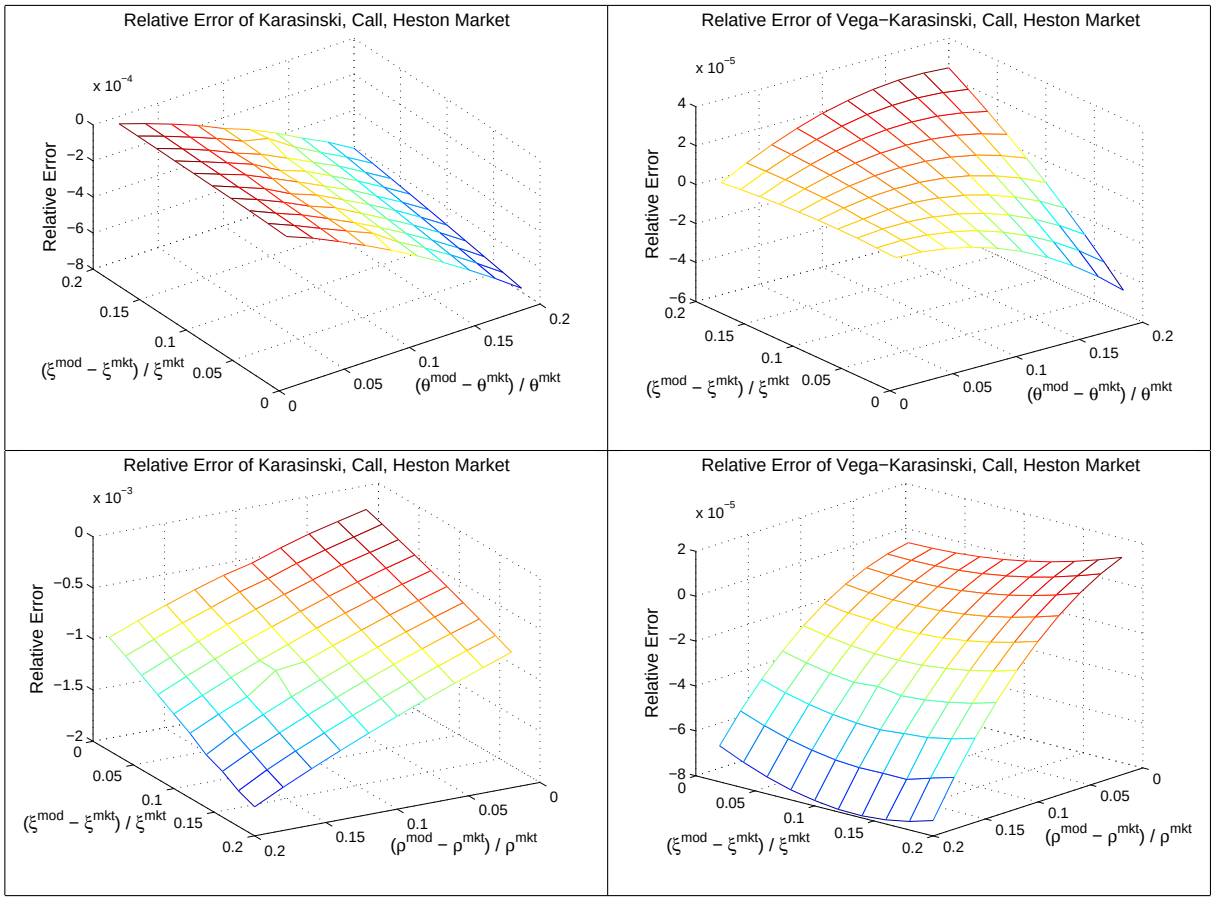

FIGURE 7.5.7. Comparison of the Karasinski (Left) and Vega-Karasinski (Right) Price Errors with respect to model ξ, θ (Top) and ξ, ρ (Bottom) changes

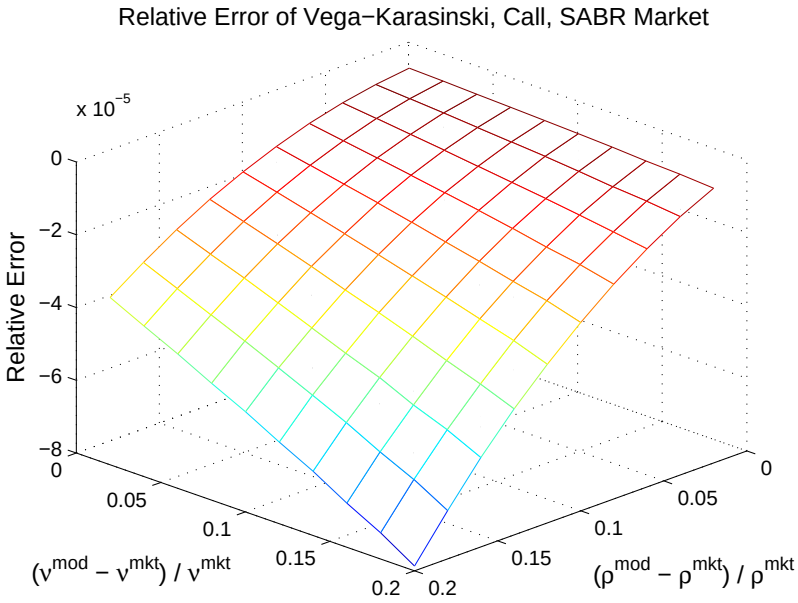

FIGURE 7.5.8. Relative Error of the Vega-Karasinski Price when SABR Market and Model Differ by Two Parameters ($N = 2$)

used in the VV method does not closely follow the SABR market whereas the SABR model used in the Vega-Karasinski method is much closer to the SABR market. From the earlier numerical analysis, we have learned that the pricing error of the Karasinski method will increase when the model parameter estimation errors increase. If we roughly view the Black-Scholes prices as some poorly calibrated SABR prices, it is reasonable to expect that the resulting pricing error of the VV method will be higher than that of the Karasinski method. Also the sensitivities of a model may be very much different from those of another model; so it is better to use a calibrated model (although the model may be not accurately calibrated) as a starting point in the Karasinski method than to use the Black-Scholes model.

With the same parameters as those for Figure (7.5.9), Figure (7.5.10) shows the comparisons between the SABR model Greek values and the Vega-Karasinski Greek values for Delta, Gamma and Vegas with respect to ρ and ν. The Vega-Karasinski Greek value are shown with much higher accuracy (10 to 10^4 times more accurate) than the SABR model Greek values. This suggests that the Karasinski method is highly accurate in computing the Greeks. Compared to the usual methods, which are the Black-Scholes and the model values, the Karasinski method appears to enjoy a clear advantage.

7.5.5. Confirmation of the Parameter-Error View - the Karasinski Method Applied to Barrier Options. When the model and market are not similar, we have argued that a hedging-costs view on the model price correction is appropriate. In the last chapter on the VV method applied to barrier options, we presented a special case where the model and the market both follow the flat vol Black-Scholes model but with a difference in the Black-Scholes implied volatility value. In this case, Figure (6.4.1) shows that the VV price is far better than the P-VV prices; this result is in sharp contrast with the other test cases in the same chapter where the model and the market are assumed to be different (e.g. the market is the Heston model while the model is the Black-Scholes model). When the market and the model are different, the P-VV prices have been shown to be better than the VV price unless the no touch probability is very close to one. If the the hedging-costs view is proper for the Karasinski method, following the same analysis presented in the last chapter, a P factor must be used for the Karasinski method as well:

$$V^{Kar} = V^{mod}\Big|_{x^{mod}} + P\sum_k \omega_k \left(U_k^{mkt} - U_k^{mod}\Big|_{x^{mod}}\right) \tag{7.5.1}$$

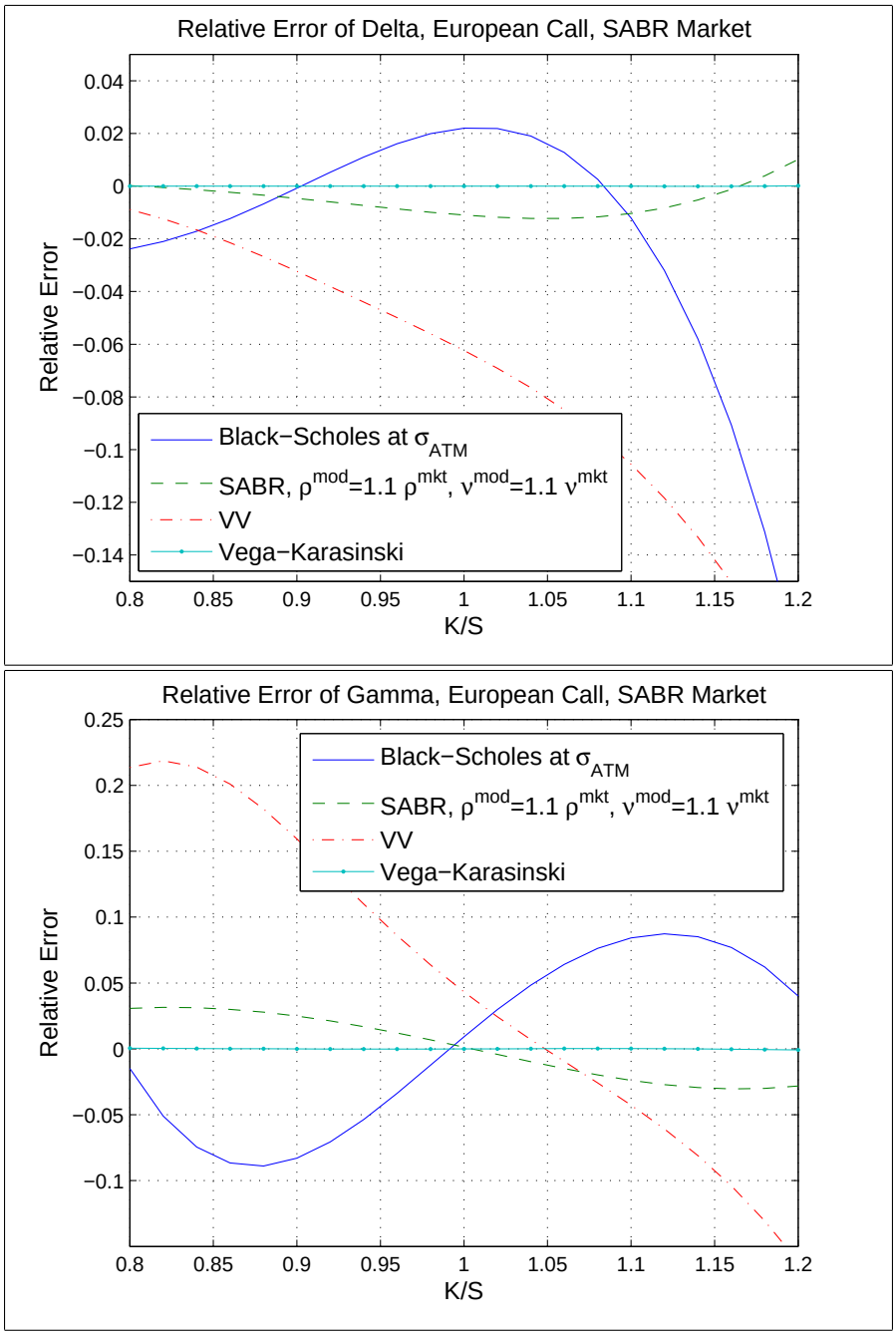

FIGURE 7.5.9. Relative Errors of Delta and Gamma Computed by the Black-Scholes, SABR, VV and Vega-Karasinski methods under the SABR Market ($N = 2$)

Otherwise, if the parameter-error view is correct, the P factor is not needed. When the model and the market are assumed to be close (as both follow the Black-Scholes model), the VV price is one case of the Karasinski price. This suggests that the parameter-error view makes sense for the Karasinski method when the model and the market are close, and the hedging-costs view makes sense when the model and the market are different.

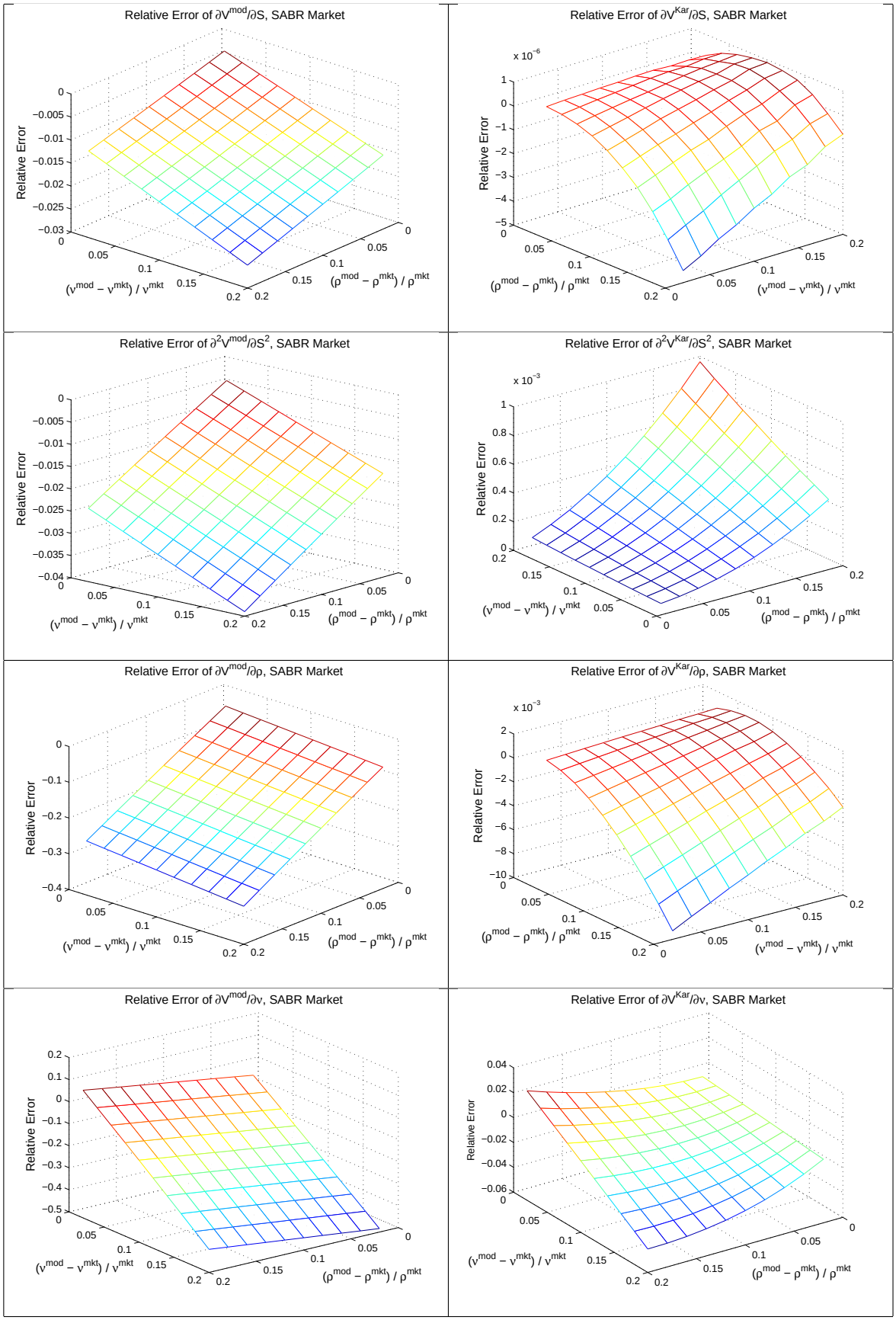

FIGURE 7.5.10. Comparison of the Model (Left) and Vega-Karasinski (Right) Greek Values under the SABR Market ($N = 2$)

Assuming the same parameters as those of Figure (6.4.4), Figure (7.5.11) presents a Karasinski test based on the model and the market both following the Heston model with the model having three wrong model parameters ($N = 3$). The model in the Karasinski method is assumed to have the Heston parameters with 10% estimation errors, $\theta^{mod} = 1.1\theta^{mkt}$, $\xi^{mod} = 1.1\xi^{mkt}$ and $\rho^{mod} = 1.1\rho^{mkt}$. Figure (7.5.11) compares the prices of the $P_{NoTouch}$-VV, P_{ex}-VV, Heston model and Vega-Karasinski prices. The Vega-Karasinski price is far better in error behaviour across the entire displayed strike range. In Figure (7.5.11), the $P_{NoTouch}$-corrected Vega-Karasinski price as in (7.5.1) shows a significantly larger error than the unmodified Vega-Karasinski price; the error of unmodified Vega-Karasinski price appears to be zero as far as our eyes can tell whereas the $P_{NoTouch}$-corrected Vega-Karasinski price has an error of around 4%. Along with the earlier test of both the model and the market following the Black-Scholes model in the last chapter, this Heston-based test again confirms that the hedging-costs view is not appropriate for the Karasinski method; the parameter-error view is more appropriate for the Karasinski method.

Figures (7.5.12) and (7.5.13) display the relative errors of the Vega-Karasinski method w.r.t. the Heston parameter value changes. It can be seen that the Vega-Karasinski method is highly accurate in pricing barrier options. As the model parameter estimation errors increase, the error of the Vega-Karasinski price will increase. However, even with 20% model parameter estimation errors, the relative error of the Karasinski price stays in the order of 10^{-4}.

7.5.6. Fast Linear Model Calibration plus the Karasinski Method for Difficult-to-Calibrate but Easy-to-Compute Models. The performance of first-order and second-order fast linear calibration based on (7.4.2), (7.4.4) and (7.4.3) is shown in Figures (7.5.14) and (7.5.15). The parameters for these figures are the same as those for the bottom plot of Figure (7.5.2).

The top plots of Figures (7.5.14) and (7.5.15) show the relative error of the calibrated ρ^{mod} and ν^{mod} based on the FLF calibration scheme. Both plots show that the calibrated parameters are closer to the market values after the calibration. For example, when ρ^{mod} and ν^{mod} before calibration both have 20% relative errors, the calibrated ρ^{mod} has the relative error of 4% and the calibrated ν^{mod} has the relative error of 1%. When the relative errors in the model parameters are within a few percent, we have shown that the

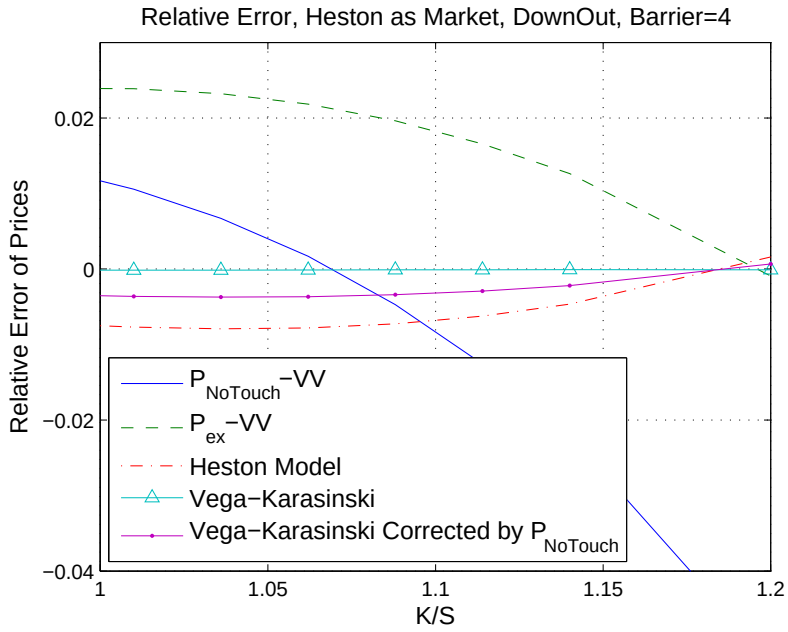

FIGURE 7.5.11. Relative Error of the P-VV, Heston Model, and Vega-Karasinski Prices w.r.t. Strike under Heston Market $N = 3$

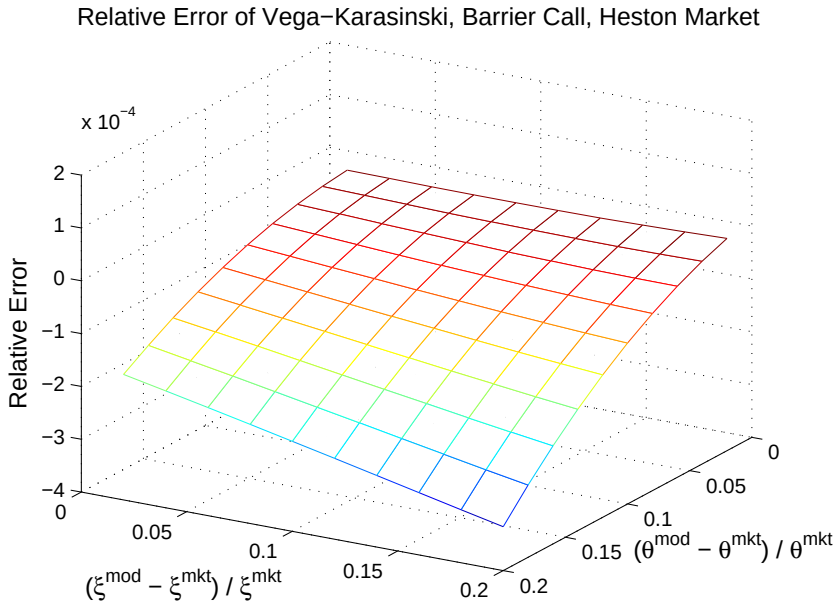

FIGURE 7.5.12. Relative Error of the Vega-Karasinski Barrier Call Price w.r.t. ξ and θ under Heston Market ($N = 3$)

Karasinski price is highly accurate. This suggests that the FLF scheme can be used with the Karasinski method for fast and accurate pricing: the FLF scheme is used to bring the model parameters close to the market values although they may be not very highly

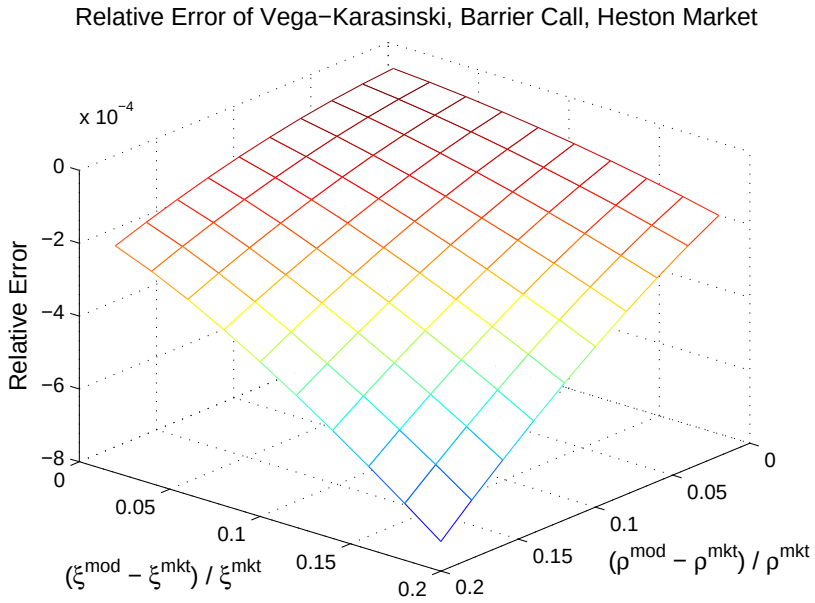

FIGURE 7.5.13. Relative Error of the Vega-Karasinski Barrier Call Price w.r.t. ξ and ρ under Heston Market ($N = 3$)

accurate; then the Karasinski method corrects the error in the model price and achieves high accuracy in pricing.

The bottom plots of Figures (7.5.14) and (7.5.15) show the relative error of the calibrated ρ^{mod} and ν^{mod} based on the FLQ calibration scheme using the combination of (7.4.2), (7.4.4) and (7.4.3). These plots show that the FLQ calibration scheme has a superb accuracy in the order of 10^{-3} when the model parameter estimation errors are within 20%. Compared to the first-order linear scheme, the second-order linear calibration scheme incurs very little extra computational costs by solving two systems of linear equations, hence the FLQ scheme is recommended over the FLF scheme.

With the same parameters as in Figures (7.5.14) and (7.5.15), Figure (7.5.16) displays the accuracy of the FLQ scheme w.r.t. a much wider range of pre-calibration model parameter errors. These plots of Figure (7.5.16) show that the second-order linear calibration scheme can provide significant model parameter estimation accuracy improvement even when the model parameters are in 80% relative errors with the market.

The linear model calibration scheme FLF and FLQ coupled with the Karasinski method may be particularly useful for the category of difficult-to-calibrate but easy-to-compute complex models. The Heston, Multi-Heston, CEV, SABR, Merton Jump models, which all have analytical solutions for vanillas, are excellent candidates that can benefit from the

combination of the linear calibration and the Karasinski method. Both the linear calibration schemes (FLF and FLQ) and the Karasinski method are extremely fast with these model. When the number of dimensions increases, a complex model will become increasingly difficult to calibrate, but a calibration using the FLF or FLQ scheme will still be very easy and fast; so these schemes and the highly accurate Karasinski method complement each other and they can offer great accuracy in pricing with very small computational costs.

7.6. Summary of Key Findings

(1) In analysing the Karasinski and VV methods, a parameter-error view is appropriate for the case when the model and the market are close but with slightly different parameter values whereas the hedging-costs view is appropriate for the case when the model and the market are different. Because the Karasinski method is intended for the case when the model and market are close, the parameter-error view is appropriate for the Karasinski method; because of the difference between the Black-Scholes model and the real market, the hedging-costs view is appropriate for the VV method. The parameter-error view for the Karasinski method is supported by the two tests with barrier options. (2) As a result of the parameter-error view of the Karasinski method, the Vanna terms need not to be balanced for computing the weights ω_k in the Karasinski method when the model follows the market. (3) The Karasinski method is a more accurate pricing method than the VV method if the model used by the Karasinski method is reasonably calibrated; our tests show that a 10% model parameter estimation error still makes the Karasinski prices highly accurate and better than the alternatives used in our tests. The closer the parameters of the model to the market values, the better the error of the Karasinski price. (4) Although the VV method can be viewed as a special case of the Karasinski method, they are intended for different applications. The VV method is used when the model and the market are different, and the Karasinski method is used when the model is close to the market but having some parameter estimation errors. (5) The Karasinski method is extended to include any number of liquid vanillas L and any number of weights-determining equations M. With the extended Karasinski method, the case of $M = L$ has a better accuracy than the case of $M \neq L$. (6) If $M > L$ holds, the Vega-Karasinski method is proposed and recommended because it cancels out the first-order error whereas the error of the Karasinski

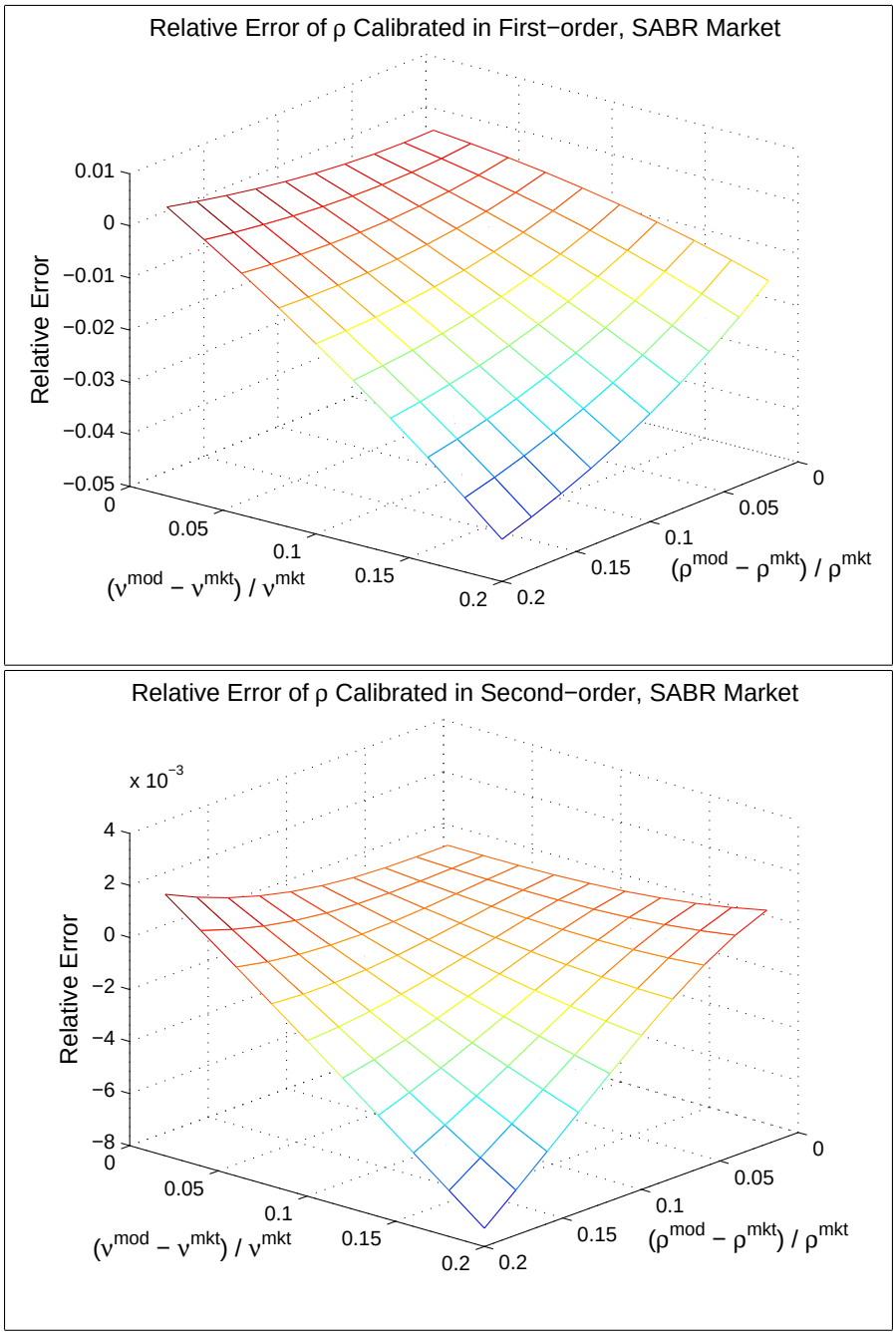

FIGURE 7.5.14. Relative Error of the ρ Calibrated in FLF (Top) and FLQ (Bottom) Calibration when SABR Market and Model Differ by Two Parameters ($N = 2$)

method with pseudo-inverse is first-order. The Vega-Karasinski method is very accurate and it is much more accurate than the Karasinski method with pseudo-inverse. In the FX market, where $M \geqslant L$ usually holds, the Vega-Karasinski method is particularly helpful. (7) If $M < L$, the use of more liquid vanillas than the number of weights-determining equations will neither significantly improve nor damage the error behaviour, hence we

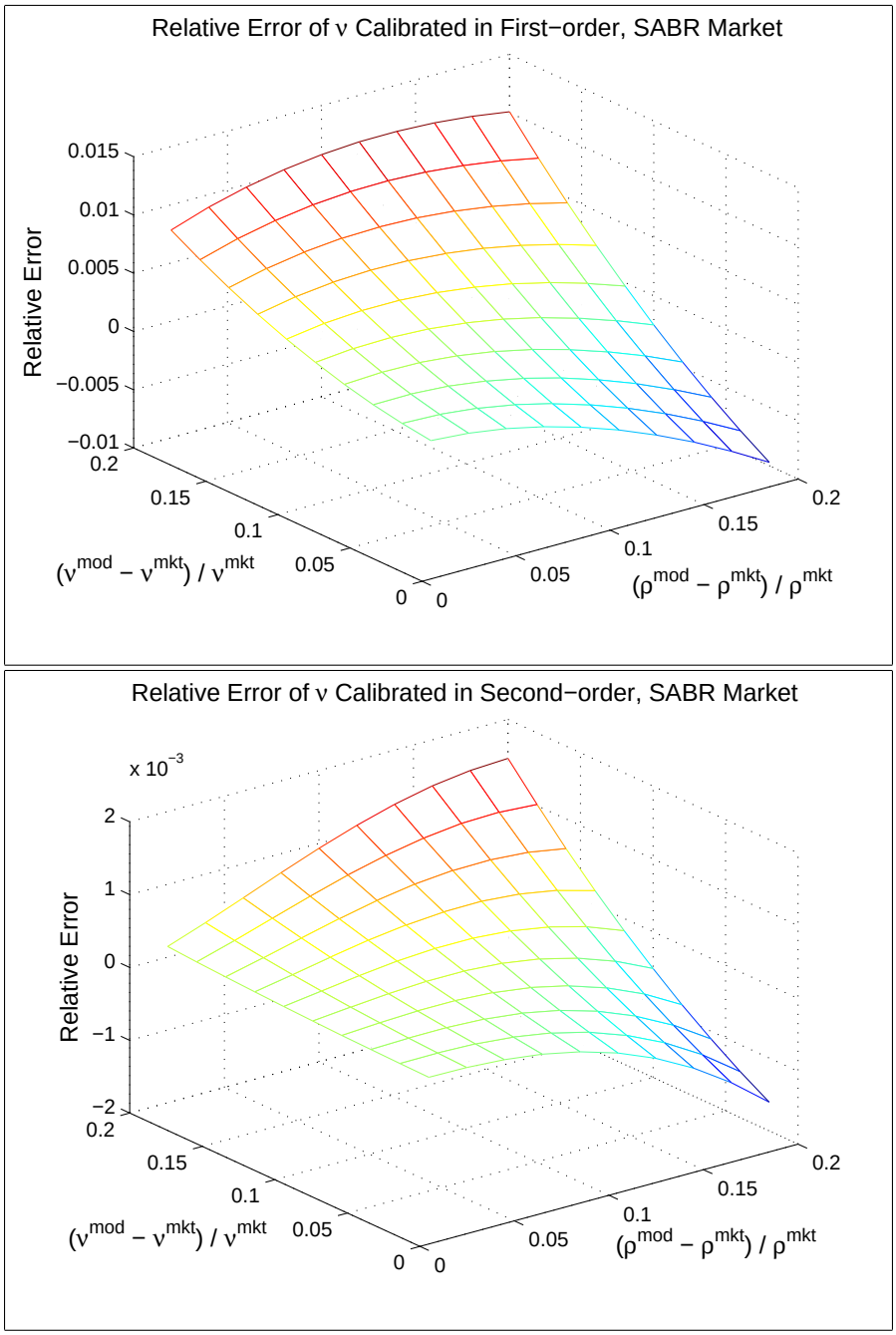

FIGURE 7.5.15. Relative Error of the ν Calibrated in FLF (Top) and FLQ (Bottom) Calibration when SABR Market and Model Differ by Two Parameters ($N = 2$)

recommend to just use M liquid vanillas to keep computational costs to a minimum. (8) For $M \geqslant L$, the Karasinski method will exactly match the market vanilla prices on the liquid strikes. (9) It is proposed that the Greeks can be computed by the Karasinski (but not the VV) method by regarding them as the target instruments. The market model Greeks computed by the Karasinski method are highly accurate. The Greeks computed

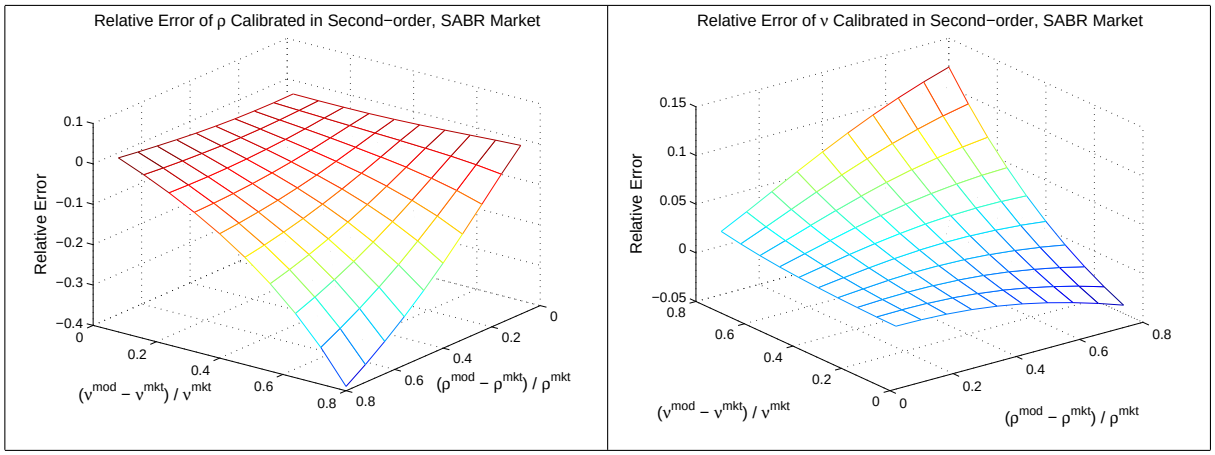

FIGURE 7.5.16. Relative Error of ρ (Left) and ν (Right) in FLQ Calibration when SABR Market and Model Differ by Two Parameters ($N = 2$)

by the VV method are far less accurate because the Black-Scholes model used in the VV method is very different from a smile-generating market model. The Karasinski method excels in computing market model Greeks in very high accuracy because the model used in the Karasinski method behaves more like the market than the flat-vol Black-Scholes model does. (10) Two fast linear market model calibration schemes (FLF and FLQ) in different accuracy are proposed to work with the Karasinski method. The idea is that since the Karasinski method does not require extremely high model parameter accuracy, a fast linear calibration scheme (in first- or second-order) can be used to bring the model parameter values close to those of the market and the remaining accuracy gap can be filled by the adoption of the Karasinski method. This way, both calibration and computation of prices will be quick and efficient. For difficult-to-calibrate but easy-to-compute models, this combination is particularly useful. The FLQ scheme is much more accurate than the FLF scheme and it requires very little extra computational cost - only solving one more system of linear equations than the FLF scheme. Therefore, the FLQ calibration scheme is the preferred scheme to work with the Karasinski method.

CHAPTER 8

Conclusions and Future Directions

8.1. Conclusions

This research work focuses on the validity and performance of the VV and Karasinski methods. We have made an attempt to conduct a comprehensive study on these methods whose merits deserve more research. The VV method has already gained a great deal of popularity as the pricing tool many traders rely upon in the FX market. Through this study, the VV method, a fast and effective but highly empirical pricing method is given the justification it needs to be better understood and explained with reasons for its success. Although the Karasinski method was proposed by Karasinski [55] as a generalisation, this method has a few interesting and powerful characteristics of its own. Potentially the Karasinski method can replace the VV method in many occasions.

In a broad sense, the VV method can be viewed as a special case of the Karasinski method. However, they are intended for different applications: the VV method is used when the model and the market are different, and the Karasinski method is used when the model is close to the market but having some parameter estimation errors. In analysing the Karasinski and VV methods, we distinguish between two views: a parameter-error view is appropriate for the case when the model and the market are close but with slightly different parameter values whereas the hedging-costs view is appropriate for the case when the model and the market are different. Because the Karasinski method is intended for the case when the model and market are close, the parameter-error view is appropriate for the Karasinski method; because of the difference between the Black-Scholes model and the real market, the hedging-costs view is appropriate for the VV method. A major difference between these two views is that the Vanna terms need not to be balanced for computing the weights ω_k in the parameter-error view whereas the hedging-costs view will require the inclusion of the balance of Vanna terms.

The VV price is found to be an excellent alternative to the Heston, SABR, CEV, Merton Jump prices for vanilla options under a wide range of market parameter assumptions; the

VV method requires no calibration, hence having a significant advantage over these model in computing vanillas. While each of these models is only suitable for certain types of smiles, the VV method can always match the market vols on the three liquid strikes, making the VV method far more flexible in modelling a smile than any of these models. The RRBF-VV price is found to be the same as the VV price; the Traders' Rule price is explained as an approximation to the RRBF-VV price. It has been found that it is necessary to use three hedging vanillas U_k in the VV formula; using one or two hedging vanillas will see a substantial increase of error and will not fit the market vols on all three liquid strikes. While perfect estimates of market vanilla prices on the liquid strikes are given by the VV method, on other strikes there are underestimates and overestimates. In an example, it is also shown that the VV weights may be excellent hedging strategies.

For exotics, there is no guarantee that the VV method will produce a perfect match to a market price. The VV prices of the path dependent exotic options are much less close to the Heston prices than the VV prices of the path independent exotic options such as digitals; the increased errors are explained by the unbalanced Vanna terms such as $d\overline{S}d\sigma_{ATM}$, $dS_{max}d\sigma_{ATM}$ and $dS_{min}d\sigma_{ATM}$ in the portfolio growth $dV^{BS} - \Delta dS - \sum_k \omega_k dU_k^{BS}$. The Smile-VV method is proposed for exotic options; the Smile-VV method has been demonstrated to perform much better than the VV method for Asian options.

The P-VV method is explained with the two issues identified with the unmodified VV method for the down-out barrier options: the inconsistency of the VV price at t_h and overhedging at $t < t_h$. From a view of solving the second issue, the $P_{NoTouch}$-VV method is found to have better underlying reason than the P_{ex}-VV method, although both P factors solve first issue completely. Although the P-VV price may be a big improvement over the VV price when the no touch probability is small, because the P-VV price is in between the Black-Scholes price and the unmodified VV price, any further improvement on error is limited by how close the Black-Scholes price is to the market price in this case. The V-VV method is proposed to help to interpolate prices between these strikes if the market barrier prices on the liquid strikes have been estimated somehow with certain accuracy. It is found that the estimation of $P_{NoTouch}$ and P_{ex} with the Black-Scholes model using σ_{ATM} is good enough for normal market parameters. We have explored the issue whether there are useful alternatives to σ_{ATM} as the model vol for the P-VV price computation, and have proposed two possible choices $\sigma^{VV}(K)$ and σ_{AVG}. The overall trend of the P-VV prices

can be summarised as that in the interpolation range of the strike, the $P_{NoTouch}$-VV price usually outperforms the P_{ex}-VV price; in the extrapolation range of the strike, the P_{ex}-VV price appears to have an advantage over the $P_{NoTouch}$-VV price. the Smile-VV method also appears to be roughly as good a method for barrier options as the P-VV method.

We have found that the Karasinski method is much more accurate than the VV method if the model used by the Karasinski method is reasonably calibrated (e.g. a 10% model parameter estimation error still makes the Karasinski prices highly accurate). The closer the parameters of the model to the market values, the better the error of the Karasinski price. With the multi-dimensional Karasinski method, the case of $M = L$ has the better accuracy than the case of $M \neq L$, where L is the number of liquid vanillas and M is the number of weights-determining equations.

For the case of $M > L$, we propose to use the Vega-Karasinski method because this method cancels out the first-order error, making the error behaviour second-order, which is better than the first-order error convergence of the Karasinski method with a pseudo-inverse. We have pointed out that in the FX market, where $M \geqslant L$ usually holds, the Vega-Karasinski method is particularly helpful. If $M < L$, we recommend to just use M liquid vanillas to keep computational costs down because using more liquid vanillas than M offers no significant increased benefits. For $M \geqslant L$, the Karasinski method will exactly match the market vanilla prices on the liquid strikes. The Karasinski method has been shown to be highly accurate for pricing barrier options.

We propose to compute the Karasinski Greeks by regarding the Greeks as the target instruments in the Karasinski formula. The market model Greeks computed by the Karasinski method are very accurate. The Greeks computed by the VV method are not accurate. The Karasinski method excels in computing market model Greeks in very high accuracy because the model used in the Karasinski method behaves more like the market than the flat-vol Black-Scholes model does.

Through the analysis and numerical tests in this thesis, the Karasinski method has been demonstrated with superior performance in computing option prices and Greeks compared to the VV method. The tradeoff of the Karasinski method is that it requires a calibrated model, although the calibration can be imperfect, whereas the VV method does not need calibration at all. Since frequent but imperfect calibration is very common among the

trading systems, this research work reveals that the existing calibration can be taken advantage of through the Karasinski method.

We have also proposed two fast linear market model calibration schemes (FLF and FLQ) to work with the Karasinski method. The logic is that since the Karasinski method does not require very high model parameter accuracy, a fast linear calibration may be used to bring the model parameter values close to those of the market and the remaining price error may be corrected by the Karasinski method. This combination is quick and efficient, and can be particularly useful for difficult-to-calibrate but easy-to-compute models. The FLQ calibration scheme can be a highly accurate calibration scheme in its own right.

This thesis also covers a careful exploration on the choices of finite difference schemes as well as the Multi-Heston model in the early chapters. The effects of the chosen schemes, Crank Nicolson, Douglas scheme, ADI time marching, Rannacher start-up, Giles extension to Rannacher start-up, Craig-Sneyd correction, payoff smoothing, non-uniform grids, and mesh-adaptive upwinding are studied. An informed combination of these schemes is fast and efficient in computing a wide range of options with non-smooth payoffs to the second-order precision. With the Fundamental Transform method (see Lewis [65] and Shaw [89]), the Multi-Heston model is studied to highlight that a slow-to-calibrate but fast-to-compute model may be a good candidate for the Karasinski method. The studies of these topics help qualify the Karasinski method as a highly practical method capable of dealing with a wide range of the trading needs.

Overall, this thesis has made a detailed inquiry into the VV and Karasinski methods. With the increased understanding about these powerful methods, plus the suitable numerical tools to make their implementations easy, it is hoped that the future use of these methods will benefit from this work.

Future Directions

A major followup of this line of discussion is to extend the VV and Karasinski methods from a largely FX method to a tool in other markets. Take the equity option market for example. There are many more liquid strikes in the equity option market whereas there are just several liquid strikes in the FX market. However, it may also be possible to simplify computation by using only the strikes of the hedging instruments around the strike of

interest. Overall, there are no significant obstacles preventing the VV and Karasinski methods from being applied in other markets.

Table of Conventions for the VV and Karasinski Chapters

FX:	Foreign Exchange	$P_{NoTouch}$:	No touch risk neutral probability
V:	The price of an option	T:	Maturity
U:	The price of a vanilla option	τ:	Time to maturity
BS:	Black-Scholes	T_{ex}:	Expected early exercise time
mkt:	Market	P_{ex}:	T_{ex}/T
mod:	Model	t_h:	The time when the barrier is hit
VV:	Vanna-Volga	S:	Spot
Vega:	$\frac{\partial V^{BS}}{\partial \sigma}$	Δ:	Delta, $\frac{\partial V^{BS}}{\partial S}$
Volga:	$\frac{\partial^2 V^{BS}}{\partial \sigma^2}$	r_d:	Domestic interest rate
Vanna:	$\frac{\partial^2 V^{BS}}{\partial \sigma \partial S}$	r_f:	Foreign interest rate
ω_k:	Hedging weights	$\mathcal{N}$:	Normal cumulative function
Kar:	Karasinski	*vol pillars*:	See K_k
ATM:	At-the-money	U_1:	The liquid vanilla 25Δ put
GC:	Gram-Charlier	U_2	The liquid vanilla ATM call
$\widehat{V}^{mkt}$:	Esitimated price of V^{mkt}	U_3:	The liquid vanilla 25Δ call
TR:	Traders' Rule	U_{ATM}:	U_2
RR:	Risk Reversal based on vols	K_k:	Liquid strike for U_k
$\widetilde{RR}$	Risk Reversal based on prices	K:	Strike of the target option
BF:	Butterfly based on vols	FLF:	"Fast linear first-order" calibration
$\widetilde{BF}$:	Butterfly based on prices	FLQ:	"Fast linear quadratic" calibration

Bibliography

[1] Abken, P., Madan, D., and Ramamurtie, S. 1996. Estimation of risk neutral and statistical densities by Hermite polynomial approximation: with application to Eurodollar futures options. *Discussion Paper, Federal Reserve Bank, Atlanta*.

[2] Adi, B. and Greville, T. 2003. *Generalized Inverses*. Springer-Verlag.

[3] Airoldi, M. 2005. A moment expansion approach to option pricing. *Quantitative Finance 5*, 89–104.

[4] Backus, D., Foresi, S., Li, K., and Wu, L. 2004. Accounting for biases in Black-Scholes. *"http://papers.ssrn.com/sol3/papers.cfm?abstract_id=585623"*.

[5] Balland, P. 2006. Forward smile. *Presentation on ICBI Global Derivatives, Paris, "http://www.scicomp.com/news/events"*.

[6] Bjork, T. 2004. *Arbitrage Theory in Continuous Time*. Oxford University Press Second Edition, Oxford.

[7] Black, F. and Karasinski, P. 1991. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal, July-August*, 52–59.

[8] Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 659–683.

[9] Bossens, F., Rayée, G., Skantzos, N., and Deelstra, G. 2009. Vanna-Volga methods applied to FX derivatives: from theory to market practice. *Social Science Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1380063*.

[10] Boyd, S. Least-norm solutions of undetermined equations, Course Notes Stanford EE Department. *"http://see.stanford.edu/materials/lsoeldsee263/08-min-norm.pdf"*.

[11] Breeden, D. and Litzenberger, R. 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business 51*, 621–652.

[12] Brenner, M. and Eom, Y. 1997. No-arbitrage option pricing: New evidence on the validity of the martingale property. *"http://papers.ssrn.com/sol3/papers.cfm?*

abstract_ id=1296404".

[13] Carr, P. and Chou, A. 1997. Breaking barriers. *Risk 10*, 139–145.

[14] Carr, P. and Wu, L. 2004. Stochastic skew models for FX options. *"http://www.math.nyu.edu/research/carrp/papers/pdf/currencyov_CU1.pdf"*.

[15] Castagna, A. and Mercurio, F. 2005. Consistent pricing of FX options. *Banca IMI Report, "www.fabiomercurio.it/consistentfxsmile2b.pdf"*.

[16] Castagna, A. and Mercurio, F. 2007. The Vanna-Volga method for implied volatilities. *Risk*.

[17] Charlier, C. 1905-06. Über das Fehlergesetz. *Ark. Math. Astr. och Phys. 2 8*, 1–9.

[18] Chawla, M., Al-Zanaidi, M., and Evans, D. 1999. Generalized trapezoidal formulas for parabolic equations. *International Journal of Computer Mathematics 70:3*, 429–443.

[19] Chebyshev, P. 1890. Sur deux théorèmes relatifs aux probabilités. *Acta Math. 14*, 305–315.

[20] Christoffersen, P., Heston, S., and Jacobs, K. 2007. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *SSRN: "http://ssrn.com/abstract=961037"*.

[21] Conte, S. and De Boor, C. 1972. *Elementary Numerical Analysis*. McGraw-Hill, New York.

[22] Conze, A. and Viswanathan, R. 1991. Path dependent options: The case of lookback options. *Journal of Finance S*, 1893–1907.

[23] Corrado, C. 2007. The hidden martingale restriction in Gram-Charlier option prices. *The Journal of Futures Markets 27*, 517–534.

[24] Corrado, C. and Su, T. 1996. Skewness and kurtos is in S&P 500 index returns implied by option prices. *J. Financial Research 19*, 175–192.

[25] Corrado, C. and Su, T. 1997a. Implied volatility skews and stock index skewness and kurtosis implied by S&P 500 index option prices. *Journal of Derivatives 4*, 8–19.

[26] Corrado, C. and Su, T. 1997b. Implied volatility skews and stock return skewness and kurtos is implied by stock option prices. *The European Journal of Finance 3*, 73–85.

[27] Craig, I. and Sneyd, A. 1988. An alternating-direction implicit scheme for parabolic equations with mixed derivatives. *Comput. Math. Applic 16*, 341–350.

[28] Cramer, H. 1957. *Mathematical Methods of Statistics*. Princeton University Press,

Princeton.

[29] Derman, E. and Kani, I. 1994. The volatility smile and its implied tree, Quantitative strategies research notes, Goldman Sachs. *"http://www.ederman.com/new/docs/gs-volatility_smile.pdf"*.

[30] D'yakonov, E. 1964. Difference schemes of second order accuracy with a splitting operator for parabolic equations without mixed partial derivatives. *Zh. Vychisl. Mat. i Mat. Fiz. 4*, 935.

[31] Edgeworth, F. 1905. The law of error. *Cambridge Philos. Soc. 20*, 36–66 and 113–141.

[32] Elder, J. 2002. Hedging strategies for financial derivatives. Ph.D. thesis, University of Oxford.

[33] Falloon, P. Option prices in Merton's jump diffusion model. *"http://demonstrations.wolfram.com/OptionPricesInMertonsJumpDiffusionModel/"*.

[34] Feller, W. 1966. *An Introduction to Probability Theory and its Applications, Vol II.* Wiley.

[35] Fried, I. and Malkus, D. 1975. Finite element mass matrix lumping by numerical integration with no convergence rate loss. *Int. J. Solids Structures 11*, 461–466.

[36] Gatheral, J. 2006. *The Volatility Surface - A Practitioner's Guide.* Wiley Finance, Hoboken, New Jersey.

[37] Giles, M. and Carter, R. 2006. Convergence analysis of Crank-Nicolson and Rannacher time-marching. *Journal of Computational Finance 9(4).*

[38] Glasserman, P. 2004. *Monte Carlo Methods in Financial Engineering.* Springer, Springer-Verlag Berlin Heidelberg New York.

[39] Goldman, B., Sosin, H., and Gatto, M. 1979. Path dependent options: Buy at the low, sell at the high. *Journal of Finance S*, 1111–1127.

[40] Hagan, P., Kumar, D., Lesniewski, A., and Woodward, D. 2002. Managing smile risk. *Wilmott Magazine.*

[41] Haug, E. 1998. *Complete Guide to Option Pricing Formulas.* McGraw Hill.

[42] Heston, S. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currecy options. *The Review of Financial Studies 6*, 327–343.

[43] Heston, S. and Zhou, G. 2000. On the rate of convergence of discrete-time contingent claims. *Mathematical Finance 10*, 53–75.

[44] Hildebrand, F. 1956. *Introduction to Numerical Analysis.* McGraw-Hill.

[45] Hinton, E., Rock, T., and Zienkiewicz, O. 1976. A note on mass lumping and related processes in the finite element method. *Earthquake Engineering and Structural Dynamics 4*, 245–249.

[46] Ho, T. and Lee, S. 1986. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance 41*, 1011–1029.

[47] Horn, B. Solving over- and under-determined sets of equations, Course Notes MIT. *"http://people.csail.mit.edu/bkph/articles/Pseudo_Inverse.pdf"*.

[48] Hull, J. 2005. *Options, Futures, and Other Derivatives.* Pearson Education Fifth Edition.

[49] Hull, J. and White, A. 1990. Pricing interest-rate-derivative securities. *The Review of Financial Studies 3(4)*, 573–592.

[50] Jarrow, R. and Rudd, A. 1982. Approximate option valuation for arbitrary stochastic processes. *J. Financ. Econ. 10*, 347–369.

[51] Jiang, L. 2003. *Mathematical Modeling and Methods of Option Pricing.* Higer Education Press, Xicheng Dist. Dewai Street, No.4, Beijing.

[52] Joshi, M. 2003. *The Concepts and Practice of Mathematical Finance.* Cambridge University Press, Cambridge.

[53] Kahl, C. and Jäckel, P. 2005. Not-so-complex logarithms in the Heston model. *Wilmott Magazine*, 94–103.

[54] Kainth, D. and Saravanamuttu, N. 2004. Modelling the FX skew. *www.quarchome.org*.

[55] Karasinski, P. 2005. Mindless fitting? *Isaac Newton Institute for Mathematical Sciences, City Event "Modelling Philosophy"*.

[56] Karatzas, I. and Shreve, S. 2005. *Brownian Motion and Stochastic Calculus.* Springer Second Edition, Springer-Verlag Berlin Heidelberg New York.

[57] Kemna, A. and Vorst, A. 1990. A pricing method for options based on average asset values. *Journal of Banking and Finance 14*, 113–129.

[58] Kloeden, P. and Platen, E. 1999. *Numerical Solution of Stochastic Differential Equations.* Springer.

[59] Knight, J. and Satchell, S. 2000. *Pricing Derivatives Written on Assets with Arbitrary Skewness and Kurtosis. In Return Distributions in Finance.* Butterworth Heinemann.

[60] Kreiss, H., Thomée, V., and Widlund, O. 1970. Smoothing of initial data and rates of convergence for parabolic difference equations. *Communications on Pure and Applied Mathematics 23*, 241–259.

[61] Kwok, Y. 2008. *Mathematical Models of Financial Derivatives.* Spinger Finance.

[62] Lax, P. and Richtmyer, R. 1956. Survey of the stability of linear finite difference equations. *Comm. Pure Appl. Math. 12.*

[63] Lee, R. 2004. *Chapter "Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation" in Recent Advances in Applied Probability.* Springer.

[64] Lehalle, C. 2006. Asian option pricing (Matlab). *"http://en.literateprograms.org"*.

[65] Lewis, A. 2000. *Option Valuation Under Stochastic Volatility.* Finance Press.

[66] Lipton, A. and McGhee, W. 2002. Universal barriers. *Risk*.

[67] Longstaff, F. 1995. Option pricing and the martingale restriction. *Rev. Financ. Stud. 8*, 1091–1124.

[68] Madan, D. and Milne, F. 1994. Contingent claims valued and hedged by pricing and investing in a basis. *Math. Finance 4*, 223–245.

[69] Douglas Jr, J. and Rachford Jr, H. 1956. On the numerical solution of the heat conduction problems in two and three variables. *Trans. Amer. Math. Soc. 82*, 421.

[70] Merton, R. 1973. Theory of rational option pricing. *Bell Journal of Economics & Management 4*, 141–183.

[71] Merton, R. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*, 125–144.

[72] Mikhailov, S. and Nögel, U. 2003. Heston' s stochastic volatility model implementation, calibration and some extensions. *Wilmott*, 74–94.

[73] Mitchell, A. and Griffiths, D. 1980. *The Finite Difference Method in Partial Differential Equations.* John Wiley & Sons.

[74] Moore, E. 1920. On the reciprocal of the general algebraic matrix. *Bulletin of the American Mathematical Society 26*, 394–395.

[75] Morton, K. and Mayers, D. 2004. *Numerical Solution of Partial Differential*

Equations. Cambridge University Press.

[76] Musiela, M. and Rutkowski, M. 2005. *Martingale Methods in Financial Modelling*. Springer Second Edition, Springer-Verlag Berlin Heidelberg New York.

[77] Oksendal, B. 2003. *Stochastic Differential Equations : An Introduction with Applications*. Springer-Verlag Sixth Edition, Berlin Heidelberg New York.

[78] Papoulis, A. 1984. *Probability, Random Variables, and Stochastic Processes, 2nd ed.* McGraw-Hill, New York:.

[79] Peaceman, D. and Rachford Jr, H. 1955. The numerical solution of parabolic and elliptic differential equations. *J. Soc. Indust. Appl. Math 3*, 28.

[80] Pelsser, A. 2000. *Efficient Methods for Valuing Interest Rate Derivatives*. Springer, Springer-Verlag Berlin Heidelberg New York.

[81] Penrose, R. 1955. A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society 51*, 406–413.

[82] Polishchuk, A. and Carr, P. 2009. Review of Vanna-Volga and its limitations. *Volatility Trends in Equity and FX Markets and Its Use in Modeling, Bloomberg, London, 17th April*.

[83] Pooley, D., Vetzal, K., and Forsyth, P. 2003. Convergence remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance 6*, 25–40.

[84] Potters, M., Cont, R., and Bouchaud, J. 1998. Financial markets as adaptive systems. *Europhys. Lett.*, 239–244.

[85] Rannacher, R. 1984. Finite element solution of diffussion problems with irregular data. *Numerische Mathematik 43*, 309–327.

[86] Reiner, E. and Rubinstein, M. 1991. Breaking down the barriers. *Risk 4*, 28–35.

[87] Richtmyer, R. and Morton, K. 1967. *Difference Methods for Initial-Value Problems*, 2nd ed. Interscience Publishers.

[88] Runchal, A. 1972. Convergence and accuracy of three finite difference schemes for a two-dimensional conduction and convection problem. *International Journal of Numerical Methods in Engineering 4*, 541–550.

[89] Shaw, W. Stochastic volatility, models of Heston type. *Lecture Notes "http://www.mth.kcl.ac.uk/~shaww/web_page/papers/StoVolLecture.pdf"*.

[90] Shaw, W. 1998. *Modelling Financial Derivatives with Mathematica*. Cambridge University Press.

[91] Shreve, S. 2003. *Stochastic Calculus for Finace II: Continuous-Time Model.* Springer, Springer-Verlag Berlin Heidelberg New York.

[92] Smith, G. 1985. *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd ed. Oxford : Clarendon Press.

[93] Spalding, D. 1972. A novel finite difference formulation for differential expressions involving both first and second derivatives. *International Journal of Numerical Methods in Engineering 4*, 551–559.

[94] Tavella, D. and Randall, C. 2000. *Pricing Financial Instruments: The Finite Difference Method.* Wiley, The Atrium, Southern Gate, Chichester, West Sussex, England.

[95] Thomée, V. and Wahlbin, L. 1974. Convergence rates of parabolic difference schemes for non-smooth data. *Mathematics of Computation 28(125)*, 1–13.

[96] Vasicek, O. 1977. An equilibrium characterisation of the term structure. *Journal of Financial Economics 5*, 177–188.

[97] Vecer, J. 2001. A new PDE approach for pricing arithmetic average Asian options. *Journal of Computational Finance 4.*

[98] Vecer, J. 2002. Unified Asian pricing. *Risk 15.*

[99] Wilmott, P. 2004a. *Paul Wilmott on Quantitative Finance I.* Wiley, England.

[100] Wilmott, P. 2004b. *Paul Wilmott on Quantitative Finance II.* Wiley, England.

[101] Wilmott, P., Howison, S., and Dewynne, J. 1995. *The Mathematics of Financial Derivatives: A Student Introduction.* Cambridge University Press, Cambridge.

[102] Wu, L. Implied volatility surface. *"http://faculty.baruch.cuny.edu/lwu/9797/Lec8.pdf"*.

[103] Wystup, U. Vanna-Volga pricing. *"http://www.mathfinance.de/wystup/papers/wystup_vannavolga_eqf.pdf"*.

[104] Wystup, U. 2006. FX options and structured products. *Wiley.*