Calibration of the Heston Model with Application in Derivative Pricing and Hedging

Chen Bin

December 18, 2007

Executive Summary

Because of the complexity of the modern financial derivatives, like option contracts, practitioners heavily rely on mathematical models to price the derivatives. It is known that the classical Black-Scholes option pricing model is not capable of pricing without a significant bias. Numerous model extensions have been introduced in recent years and in this thesis we discuss several of them. To improve the understanding of the dynamics of the future uncertainty of asset prices, which is measured in volatility, and its impact on risk-free 'hedged' portfolios, the stochastic volatility models have emerged in the last decade of the previous century. Within the big family of Stochastic volatility models, the Heston model (one of the many stochastic volatility models) has become a new industrial standard in the domain of exotic equity derivative. Its popularity comes from the fact that the Heston model can price European options highly efficiently by means of the so-called Fast Fourier Transform (FFT) algorithm.

Advanced model requires equally sophisticated empirical implementation, in which stage the calibration problem comes in. Since the Heston model contains several undetermined parameters that need to be fitted to the present financial market data. This usually leads to an optimization problem, as calibration implies that the 'distance' between market and (Heston) mathematical model prices should be minimized. The major difficulty (the one we describe in detail in this thesis) is that the optimization problem is typically ill-posed: The commonly used methods may generate unstable parameters through time, or cannot produce a sufficient fit to the market prices. Here we propose a *novel* multilevel-structured global optimization procedure, called the *Hybrid Stochastic Approximation Search*. In this algorithm, we especially contribute with two distinct Markov processes: The first one Wang-Landau algorithm is capable of leaving local optima, in which common gradient-based optimization methods would stagnate in their convergence. At the same time, this algorithm is not as costly as common global optimization algorithms are; Secondly, we propose to

include a technique called 'Partial resampling' in order to reduce big fluctuations produced by the first level and steering the steps through the local dynamics of the optimization landscape.

The optimization method is called a 'Stochastic Approximation' because the Wang-Landau is a controlled Markov Chain guided by the historical samples in the parameter space. These two processes are combined in a multi-level structure where level one, based on the Wang-Landau algorithm ensures a global search, and level two, the partial resampling, increases the efficiency by steering the algorithm with problem specific information.

With the calibrated Heston model, we obtain a satisfactory low pricing error for basic 'plain vanilla' options with parameters that are, to a large extent, stationary, i.e., not depending on time. A local enhancement of calibration can be made by allowing certain parameters to be mildly dependent on time.

Next to this main contribution we considered it important to also present the typical use of these mathematical pricing models, and in particular the stochastic volatility models, in financial practice. Therefore, a chapter is reserved for the explanation of the pricing of exotic contract as well as for the hedging of a portfolio, as it occurs in practice. We focus here on the so-called 'hedge ratios' that are implied by the Heston model and derive analytical solutions for financial contracts that are considered to be volatility sensitive, like cliquets and variance swaps.

Keywords: computational finance, advanced asset price modeling, stochastic volatility model, calibration of parameters, stochastic optimization, pricing of financial derivatives

Preface

Because of the presence of the randomness and uncertainty in the payoff of financial derivatives, practitioners rely on mathematical models to understand the likelihood of and proximate the 'risk' – one cannot manage it which one cannot measure. An increasing sophistication of the financial models requires equally sophisticated methods for their empirical implementation. One of the most important implementation issues is the model calibration problem.

This thesis serves as a documentation of my work during the project which has been made at Rabobank International, from March 2007 until November 2007. The project deals with the problem of calibrating parameters in financial mathematical models. The asset price model under consideration here is the well-known Heston model. This is a very popular alternative to the classical Black-Scholes model due to a superior hedging performance. Unfortunately, there is still no universal way of obtaining the model parameters from financial market data.

In the industry, one way of calibration is the so-called 'implied' methodology for derivative pricing. Calibration boils down to solving an inverse problem: First one assumes a parametric model which governs the underlying asset price process, then one tries to find the parameters such that the difference between the prices generated by the model and the financial market prices is smallest.

The difficulty of the implied approach for model calibration is that it often gives rise to an ill-posed inverse problem: this method may generate unstable parameters through time, or it can not produce a sufficient fit to the market prices. In this thesis we aim to overcome these difficulties by developing a novel global optimization algorithm, called the *Hybrid Stochastic Approximation Search* algorithm to deal with the ill-posedness in a robust and efficient way.

To aid the reader, we finish this preface with an outline of this thesis.

Outline of the Work

Chapter 1

The first chapter describes the problem setting and the validity of underlying assumptions. Here we discuss the risk associated with the financial derivatives and formulate the calibration problem.

Chapter 2

In Chapter 2 we select the pricing model for the derivatives, based on an analysis of the problem of the so-called implied volatility in Black-Scholes models. Stochastic volatility and jumps in asset prices are essential features for mathematical models that construct models with features that are consistent with market data. Many stochastic volatility models generate, more or less, the same (implied volatility) surface. Therefore, the model of choice is the Heston model because it has a semi-analytical Fourier solution for option prices and hence derivative pricing is relatively cheap.

Chapter 3

Chapter three describes a numerically stable and efficient calibration procedure for the calibration of all parameters and for fast daily recalibration. The approach introduced is based on a least squares formulation which is solved by a multi-level hybrid stochastic optimization algorithm with also a deterministic search component.

Chapter 4

Chapter four gives a detailed description of the multilevel-structure global optimization algorithm, named the Hybrid Stochastic Approximation Search (HSAS) algorithm. We use a gaining factor (updated at each iteration) to control the first level of the search process in order to achieve broad and efficient exploration. Then we randomly fix parts of the parameters to search certain subspaces in the second level. In the end, a deterministic search is applied to find the global optimum.

Chapter 5

Numerical results for optimizing some test functions and a model calibration based on true Heston parameters is presented in Chapter 5.

Chapter 6

This chapter finally presents several applications of the Heston model for pricing and managing some exotic derivative securities, like the variance swap or the cliquet option.

Delft, December 2007

Acknowledgments

I would like to thank my father and mother who gave me the freedom to take the adventure intellectually and do something useful. Though they can not understand English, but I am sure that they watch over me through the whole project and feel proud of this thesis.
I am deeply grateful to Prof. Dr. Cornelis Oosterlee (Mathematical department of Technical University of Delft) for his great help during this project, for his patience, readiness to support and face-to-face meetings. Without his continuing advice, mathematically as well as personally, this thesis would not have been possible.
Moreover, my deepest gratitude goes Prof. Dr. Alexander Verbraeck for his admirable deep knowledge of many different areas, ranging from mathematics, computer science to project management, which was a great motivation throughout this project.
A lot of thanks goes to Prof. Dr. Jan van den Berg for his continuing help and trust in my work which were essential for the success of this project. I am very grateful not only for his willingness to support this work, but also for his invaluable guidance throughout the last year.
I should not forget my colleague in Rabobank Herwald Naaktgeboren, Leszek Grzelak and Tim Dijkstra, with whom I could discuss many details of this work and they gave me many inspiring advices.
I also want to thank my old friends who endured my mental and physical absence over the last two and a half years. All of them have been very patient and understanding, and I am grateful for their willingness to keep up with much pressure on our friendships (but I promise to make it up!)
Thank you.

Contents

List of Tables

Chapter 1

Introduction

> *...how to go from probabilistic assessments to decisions. Managing risk in general, and managing financial risk in particular, is clearly a case of decision making under uncertainty.* — Riccardo Rebonato, *Global Head of Market Risk, RBS*

This chapter reviews the problem setting and current state of financial derivative modelling. A historical approach is taken here because the present problems are embedded in a historical context, and triggered by 'almost accidental industrial and technological developments' [77].

1.1 Role of Models in Financial Derivatives

The concept of financial derivative is not new. As Chance [24] pointed out, the use of derivative contracts dates back to 1700 B.C.[1]. Unlike any other contract, derivative contracts would be worthless in the absence of uncertainty. The notion of risks and uncertainty are thus central to the business of financial derivatives.

After the birth of the celebrated Black-Scholes (BS) formula, the derivative industry has grown substantially in terms of size and complexity. The increasing complexity of derivative contracts has long been linked to the parallel development of mathematical

[1] Genesis, Chapter 41. According to Joseph's advice, an Egyptian Pharaoh, anticipating seven years of feast followed by seven years of famine, executes a hedge by storing corn. Joseph is put in charge of administering the program. See 'A Chronology of Derivatives' by Don M. Chance for an extensive chronology on derivatives pricing from 1700 B.C. to 1995.

tools for efficient pricing and hedging. This contract, illustrated by the payoff function below, is an example of the structured bond the banks are selling:

$$\text{Payoff} = \left\{ \sum_{i=1}^{d} \min \left(\text{Cap}, \max \left(\frac{S_{t_i}}{S_{t_{i-1}}} - 1, \text{Floor} \right) \right) \right\}$$

It is called a global floored cliquet. The motivation for the investors to buy such contract is that it protects investors against downside risk, yet with a significant upside potential. The banks charge the client a premium for taking on the risk and for having to manage the deal in the future. Without the premium, the structure priced at 100 euro may well have been sold at, for example, 92 euro. Now it is up to the complex derivative trading desk at the bank to 'protect' an expected 2 euro profit throughout the life of the deal, until it matures. To do this, the desk will go into the vanilla market and hedge the risks at the cost of 2-10 euro.
Typically, hedging with options is done in two ways: Replication and dynamic hedging (managing) portfolio value sensitivities.
The models that developed from the family tree of Black-Scholes model, in essence, shared the common assumptions that the payoff of any tradable financial derivatives can be replicated by a portfolio of asset underlying the derivative. So that estimation of the 'trend' and randomness of the underlying factor or financial driver was not relevant to arrive at the price of the derivative product. This insight is drawn from the concept of payoff replication. The Black-Scholes model is where the idea of replication starts:

DEFINITION 1.1.1 (Arbitrage Free (Merton) [75])**.** *The issuer of a contract on the index tries to trade in the index itself such that the value of the trading portfolio is always equal to the value of the obligations arising from the contract. If this is possible, then the 'fair' price of the contract must be the cost of the associated replication strategy and since there is no risk involved in executing the strategy, this price is independent of any risk-preference.*

In the post Black-Scholes age, the replication idea has become 'the bedrock of option pricing' (Rebonato) [77]. It is the case that when we assume the payoff can be replicated perfectly, then we don't have to worry about aversion to risk at all. However, this assumption is not true in reality. When the market reacts to new information, the implied volatility starts to fluctuating. In this kind of situations, the fair option values can not be

determined, thus a perfect replication is infeasible to achieve. As an alternative to the paradigm of replication, traders usually dynamically hedge the risk exposure of the derivative they sell or hold.

1.2 Model Users and Problem Setting

To understand the problem context, it is relevant to know about the people in the financial industry, who use the mathematical models and how they apply them. Typically, a pricing model is of interest to so-called *plain-vanilla-option traders*, to *relative-value traders* and to *complex-derivatives traders*. Plain-vanilla and relative-value traders are interested in these models because of their ability to predict how option prices move relative to the underlying, and relative to each-other, given a certain move in the underlying. For complex-derivative traders (the target users for the research in this thesis) the objective of using a model is somewhat different. The complex derivative product which they trade usually has an exotic structure and is therefore not liquid in the market (see Figure 1.1), meaning that this product is not frequently traded. Traders therefore do not have access to liquid market prices for these products, and expect to gain insight in a fair product price with the mathematical models, given the observable market data for the underlying. Since complex derivative traders also use plain vanilla derivative instruments to hedge their positions, the correct reproduction of the prices of these instruments – the *descriptive* aspect of a model – is of the highest importance.

OBJECTIVE 1. *A well-calibrated mathematical model should be able to recover the current prices of plain vanilla hedging instruments with highly satisfactory accuracy.*

1.2.1 Hedging in Practice

Probably no other aspect of derivatives trading has a bigger impact on the prices than the practice of hedging. In a Black-Scholes model, the hedge of an option is done by taking a position in 'delta' amounts of stock, where the number 'delta' represents the derivative of the option w.r.t. the asset. However, besides a delta neutral hedge, traders also have to establish a 'vega' hedge in order to cancel out the effect of a changing volatility. Once the hedging portfolio has been set up, a trader will carry out a *dynamic hedging strategy*, i.e., he/she updates the delta and vega hedge until the option's expiry.

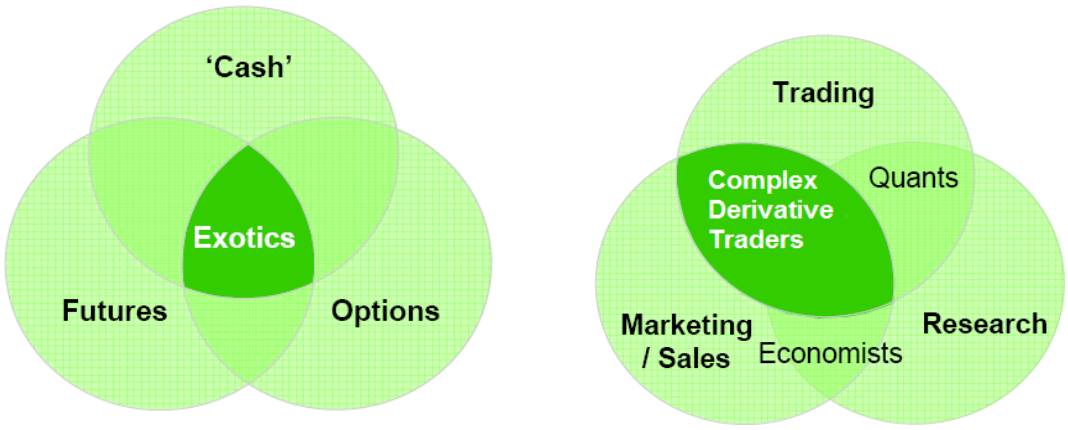

Figure 1.1: The setting of a typical trading floor schematically depicted. Left-side figure shows the typical products on the trading floor. Right-side figure displays who is the problem owner and where does it fit in.

How to Hedge Manage Risk of a Position?

Here we take a trade of a plain-vanilla instrument as an example.

◇ *August, 2006*: A trader sells a 2 year *call* option of ABN AMRO struck at 120% of initial price on a notional of $10,000,000 for an implied volatility of 14.00%. At this point in time, the stock price was 23.43 euro, and the 50 days realized volatility was 11.00%. The option was worth 1.9075 euro.

◇ *Half year later (early March, 2007)*: Barclays disclosed a takeover bid for ABN. The stock price went up to 27 euro and was expected to rise further. The option was then worth 3.9157 euro (with an implied volatility of 23%). The option trader (the 'writer' of the contract) knows that if he had not hedged his position well he would face a severe loss by selling the ABN stock at the agreed price. Therefore the trader, at the start of the option contract, already bought a certain amount of stocks (which also increased in value), to cancel out the risk related to the move of the stock. At the same time he bought a variance swap as a vega hedge position to compensate the volatility risk.

Technical Point: Vega Hedge

Take a Black-Scholes' delta hedging strategy for a call option:

$$\Delta = \frac{\partial C_{BS}(S,t,K,T,\sigma)}{\partial S} = \begin{cases} \frac{\partial C_{BS}}{\partial S} & \text{if volatility } \sigma \text{ is constant} \\ \frac{\partial C_{BS}}{\partial S} + \frac{\partial C}{\partial \sigma} \cdot \underbrace{\frac{\partial \sigma}{\partial S}}_{\text{Vega}} & \text{if } \sigma \text{ varies w.r.t underlying stock} \end{cases}$$

By controlling the vega risk, the hedge error can be reduced by 30%-40% (Derman [35]).

□

Let us now reconsider this trade some days later. At some day number 2, the model parameters determined at day number 0 will generally not price the plain-vanilla option in line with the market anymore. So, if the hedging transactions would be carried out with the 'day-0' model parameters, the model prices would not coincide with the market prices. To avoid this, traders usually *re-calibrate the mathematical model* on the basis of the new market prices, and recalculate the price of the complex instrument based on these new parameters, assuming again that they are valid until the option's expiry. If the new parameters give rise to very different delta or vega, a delta or vega re-hedge should take place.
Since the true cost of an option is linked to the cost of the replicating portfolio, the trader has to consider two sources of cost here: The cost of hedging which enters when the initial hedge is established and the delta/vega re-hedging cost encountered during the life of the option.

OBJECTIVE 2. *An important objective is to develop a calibration procedure that gives rise to stable model parameters and low predictable re-hedging costs. In particular, the calibrated model should offer stable model parameters through time.*

1.2.2 Formulation of Calibration Problems

Given the above mentioned demands on the pricing and hedging of complex derivatives, we need a stochastic model for asset and option prices such that:

1. All traded derivatives are correctly priced (arbitrage-free as in Definition 1.1.1) .
2. The model should be consistent with the market prices of plain-vanilla options.

3. Efficient calibration, pricing and sensitivity analysis in a real-time environment should be possible. (It should not cost too much computing time.)

4. Stable model parameters from day-to-day are preferred.

In this MSc Thesis, we focus on all these four criteria.

1.3 Calibration Methodology and Assumptions

The term 'calibration' indicates the task of estimating best-fitted parameters in a parametric model in comparison with a chosen observable quantity. The estimation method employed can be either a maximum likelihood estimation, a prediction error method (to be explained in Section 3.5), a Bayesian approach or any other method. The source of information consists typically of historical data or current market prices for liquid instruments.
In the financial industry, the prevalent method for pricing products with future uncertainty is the so-called 'market implied approach'. It assumes the current market prices to be as the relevant information source and a minimal prediction error between mathematical model and these market prices as the objective function. So, we start by observing present day market prices for plain-vanilla options, assuming that these prices are correct, in the sense that they embody in the most relevant information available about a stochastic model for the underlying asset (say, a stock). We set up a mathematical model of this underlying stock process, and calculate model option prices based on guessed parameters. If these computed model option prices are not correct, we did not discover the correct parameters of the underlying process. Subsequently, we therefore improve the model parameters by calibration until some pre-defined criterion on the quality of fit is met.
This procedure of fitting prices is based on the assumption that a trader, who chooses and calibrates the model, agrees with the view that the market option prices are fully consistent with the true, but a-priori unknown, mathematical process for the underlying asset Rebonato [77]. The market, on the other hand, is assumed to be *a perfect informational machine*, which absorbs all the relevant information about the unknown 'stock' process, and produces consistent option prices.
Two assumptions, on which the calibration method developed in this thesis are based, are:

1. The volatility follows a stochastic mathematical process;

2. The market has incorporated all the relevant information in the prices of European options.

We start our discussion on the validity of these assumptions with the so-called *efficient market hypothesis*(EMH), which is at the core of the informationally efficient financial market assumption.

1.3.1 The Efficient Market Hypothesis (EMH)

Strictly speaking, the EMH requires that all economic agents are fully informed and perfectly rational. 'If they could all observe the same history (of prices, economic variables, political events, etc), they would all arrive at the same statistical conclusions about the real world, and would determine prices by discounting the expected future cash flows from a security at a discount rate dependent on the undiversifiable uncertainty of the security and on their risk aversion' (Shleifer [90]). In this sense the value of the derivative embeds all the relevant information available in the market, meaning that its value is fair, and markets is said to be informationally efficient.
Obviously, 'the fully informed and rational' agents statement is too strong an assumption to be true in reality. A weaker form of market efficiency does not require all economic agents to be rational, but assumes certain investors who price securities with some imperfect information. The question is whether this will affect the market price. If we could assume that the actions of the uninformed, irrational investors are 'uncorrelated and random', their actions would cancel out and the market would agree on the same prices. However, the zero-correlation assumption is too strong. One can just look at the behavior of investors in the dot.com era, which was highly corrected. This phenomenon, however, does not imply the end of the efficient market hypothesis.
As long as there exist rational well-informed agents, as stated in Rebonato [77], who can value securities on the basis of fundamentals and trade freely, the prices will return to their fundamental values. This group of well-informed investors is called the group of pseudo-arbitrageurs. They buy the 'irrationally cheap' securities and sell the 'over priced' ones. By so doing they implicitly drive the prices back to their fundamental values. In the 'pseudo-arbitrageurs' context, excess of demand automatically creates extra supply, and

vice versa. Therefore, supply and demand will not affect the equilibrium prices and the efficient-market hypothesis would still apply.

1.3.2 Practical Recommendation for Calibration

Returning to the calibration problem, we observe a prevalent market preference for the market-price-implied-calibration approach over the use of historical time series of stock prices. This practice makes sense only if the derivative market is informationally efficient. From the discussion of the EMH, there appears to be sufficient reason (the existence of pseudo-arbitrageurs, for example) to rely on the informational efficiency of the European option prices used in model calibration. However, practitioners should remember that every model has its domain of applicability beyond which it ceases to be useful. One should always ask oneself: 'If the prices fail to reflect the best estimate of a quantity, how easy would it be for a clever trader to exploit this?'. If the answer is that the pseudo-arbitrage (i.e., the correction of the market) is simple and relatively risk-less, one can assume that the market prices do convey useful information. If, however, the correctional trades would for any reason (like the recent sub-prime mortgage problems in the beginning of August) be difficult to implement, one should be cautious and only rely on the 'best market estimate' with some reservation.

Chapter 2

Smile Consistent Option Pricing Models

> *Option pricing theory is the most successful theory not only in finance, but in all of economics.*
>
> — Steve Ross, *chaired professor at MIT*

Since the Black-Scholes ground breaking option pricing formula in 1973, a huge market of financial derivatives on a wide range of risk factors and economic quantities has developed. However, this model is not able to capture the so-cabled 'volatility smile' of the implied volatility of plain vanilla options. The origin of this problem is that in the BS model a constant volatility is assumed.
The stochastic nature of the variance level of the stock price process becomes critical if we price instruments that are very sensitive to volatility risk, such as a variance swap or a cliquet-type structure. The purpose of this chapter is to define the problem, to search in the literature for possible solutions and to motivate the choice of a specific model.

2.1 The limitation of BS Model and Implied Volatility

Black and Scholes use a bond and a stock to set up the replicating portfolio, where the bond follows a deterministic process and the interest rate r_t is assumed to be constant:

$$d\beta_t = r_t \beta dt; \tag{2.1}$$

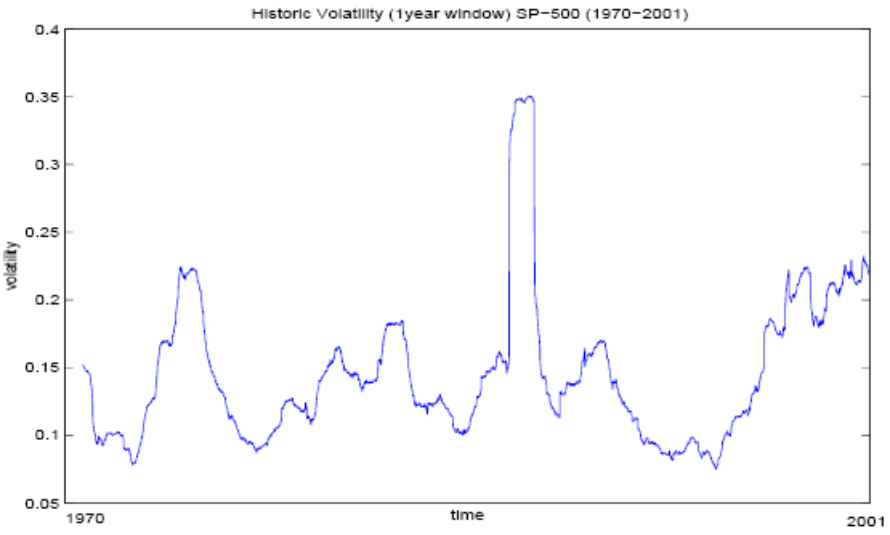

Figure 2.1: The non-deterministic behavior of the volatility estimated over time (Schoutens [85]

.

The stock price, S_t, is based on a geometric Brownian motion and $x_t := \ln S_t$ is governed by the following stochastic differential equation:

$$dx_t = (\mu - \tfrac{1}{2}\sigma^2)dt + \sigma dW_t,$$

where μ is the drift, σ the volatility. The price of a European option, V, can be obtained by solving the following Black-Scholes PDE:

$$\frac{\partial V}{\partial \tau} = \tfrac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + (r - \tfrac{1}{2}\sigma^2)\frac{\partial V}{\partial x} - rV.$$

This can be derived, for example, by using Itô's Lemma. The advantages of this model, which represents the foundation of modern finance, lie in its simplicity and clarity. However, the assumptions on the validity of this model, i.e. the log-normal return of the underlying asset and the constant volatility, are problematic as can be seen in Figure 2.1. If the Black-Scholes' assumption of constant volatility would be correct, then the *implied volatilities* of all options of the same type should be constant across strike prices. Unfortunately, this is not the case as market prices show a smile pattern in implied volatilities across strikes and maturities. The implied volatility value is the volatility

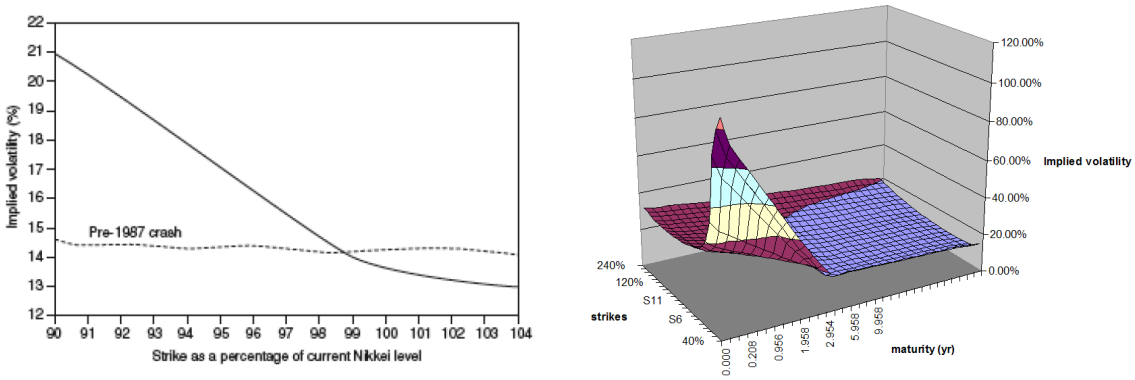

Figure 2.2: Left side: The smile is a post-crash phenomenon. Since more people buy options for protection against crashes; Right-side: The short end volatilities are highly skewed.

parameter that, when substituted in the Black-Scholes option pricing equation matches the market option price. It can be seen as the 'market's estimation of volatility'. By quoting implied volatility levels one can compare option prices across strikes and expirations.
An empirical explanation for the 'volatility smile' observed in many markets is that after the 1987 stock market crash, the market started to be more volatile than previously assumed. This behavior cannot be modeled by the assumption of normality. More precisely, the Out-of-The-Money (OTM) strikes have a higher probability to move to the At-The-Money (ATM) region and the market prices for an OTM option are therefore higher than the computed Black-Scholes model prices. Given that the volatility is the only unknown parameter in the mathematical model, the way for Black- Scholes model prices to equal market prices is to increase the volatilities for the OTM options. Consequently, implied volatilities of OTM options will be higher than implied volatilities of ATM options, see Figure 2.2.

2.1.1 Search for Smile Consistent Models

The main motivation for searching improved option pricing models is the above mentioned empirical evidence of the BS model's pricing errors across both strikes and maturity (financial practitioners called it implied volatility 'smile'). Such evidence suggests that implicit stock return distributions are 'negatively skewed with a high kurtosis' (Bakshi, Chen [7]). Guided by this insight, the search for improved models focuses on finding an

improved distributional assumption [7].
In the past two decades, the industry has developed a wide variety of alternative models that can resolve some of the limitations of the Black-Scholes model. Notable ones include:

(i) the stochastic-interest-rate option models,

(ii) the jump-diffusion model, Merton [72],

(iii) the constant-elasticity-of-variance model, Cox, Ross [30],

(iv) the stochastic-volatility models, Heston [52],

(v) the stochastic-volatility and stochastic-interest-rates model, Scott [86], and

(vi) the stochastic-volatility jump-diffusion models, Bates [9].

This list is by no means complete and the number of mathematical models is still growing. The difficulty of choosing between so many models overwhelms anyone, researchers and financial professionals alike.
Each option pricing method is based on three basic stochastic processes – one for the underlying price, one for the interest rate and one for the so-called market price of risk [6]. For each of these, there are several choices to be made. For example, the underlying price process can be either a continuous-time or a discrete-time process; With respect to the continuous-time process, it can be Markovian or not, diffusion-type or not, Poisson-type or a not, or even a model consisting of a mixture of jump and diffusion processes with or without stochastic volatility and with or without random jumps. For the interest rate process, there is a number of choices as well, but this topic is not addressed here. It seems that the search for a better option pricing model is endless, because the accurate prediction of the future behavior of stock and option prices is of high interest to many people.

Problem Statement:

From a practical point of view, a perfectly specified option pricing model is too complex. So, in the end, it is a choice among models with some imperfections, but the major task is to choose a model that

(i) is best suited for its task in the sense that it produces the lowest pricing errors, and

(ii) show a highly satisfactory hedging performance,

(iii) at a reasonable implementation cost.

REMARK. The pricing error measures the static performance of a model, whereas the hedging performance measures the dynamical properties of it [6]. However, it is not feasible to include a hedge test in this research because of lack of time. Some discussion of the results in the literature on this is presented here.

An important paper on the empirical performance analysis of alternative option pricing models is by Bakshi, Cao, Chen [6]. They use the following three criteria to judge the quality of an alternative model:

1. The consistency of the computed implied structural parameters with those in the relevant times-series data, like in the implied-volatility time series. If an option model is correctly specified, its structural parameters will be consistent with those in the observed time-series data.

2. Out-of-sample, i.e., extrapolated pricing errors give a direct measure of the model's misspecification. In particular, while a more complex model will generally lead to a better in-sample fit, it need not necessarily perform better for out-of-sample data.

3. Hedging errors measure how well a model captures the dynamical properties of option prices w.r.t. the underlying prices of the stocks.

Models considered in Bakshi [6] are the Black-Scholes (BS) model, the stochastic-interest-rate (SI) model, the stochastic-volatility (SV) model, the stochastic-volatility plus stochastic-interest-rate ($SVSI$) model, and the stochastic-volatility-jump (SVJ) model. Based on more than $38,000$ S&P500 call option prices from June 1988 to May 1991, they concluded the following:

1. Judging on the internal parameter consistencies, all models are imperfect to some extent. However, the SVJ model performs best and the BS is the worst.

2. Out-of-sample pricing errors are the highest for the BS model, and second highest for the SV models. They are lowest for the SVJ model. Overall, the stochastic volatility feature achieves the highest pricing improvement and typically reduces the BS pricing errors by 25% to 60%.

3. Adding the random *jumps* improves the fit especially for *short-term* options and including the *SI feature* enhances the pricing fit for *long-term* options. When both,

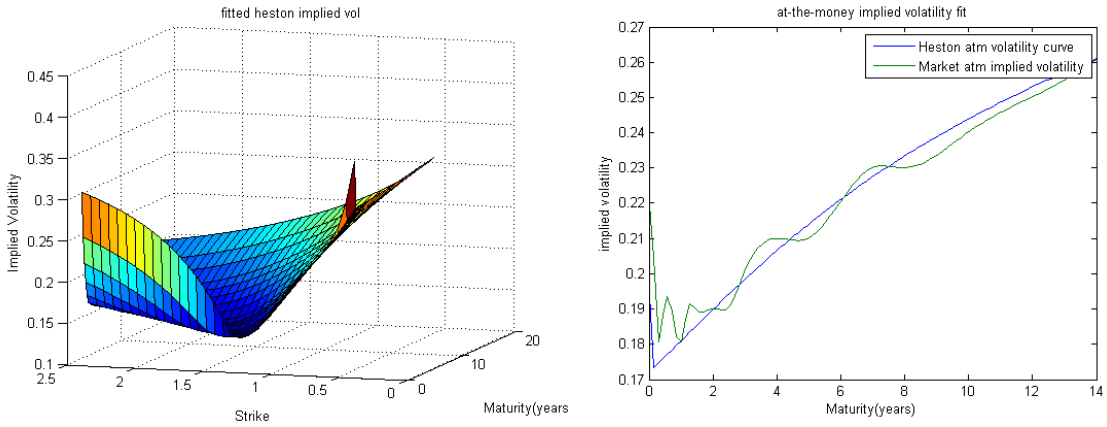

Figure 2.3: Left side: A volatility surface generated by the Heston model. Asymptotically, the Heston tends to produce a wing-shaped surface on two sides of ATM strike (at the short maturity end). Right side: The fit to term structure of ATM volatility is not good at the short end, because of the presence of the vol-of-vol term which produces the deep dip of volatility. However, the fit to long maturity volatility is satisfactory.

stochastic volatility and random jumps, are included, the pricing errors do no longer exhibit a clear systematic bias across strikes or maturities.

4. The addition of stochastic volatility to the model, reduces the hedging error by about 50% to 65% compared to the BS model (given that the hedge is re-balanced daily). Adding the jump feature does not enhance the hedging performance significantly.

The overall conclusion in [6] is that the stochastic volatility feature is most important, as it gives a significant improvement in terms of all three quality criteria. Its limitation is 'its ability to model sufficient short-term kurtosis (so its ability to price short-term options correctly)' [6]. In principle, the jump and the stochastic-volatility features improve the pricing and hedging of, respectively, the short-term and certain long-term options. However, this comes at the cost of increased model complexity. In the next section, we introduce the models that incorporate stochastic volatility and jump features.

2.2 Stochastic Volatility Jumps Models

This section describes the Heston model and its extension with jumps, called the Bates model. The specific advantages of the Heston and Bates models, among the other *stochastic volatility* models, are the existence of a fast and easily implemented semi-closed

form solution for European options. In contrast, determining accurate option prices from stochastic volatility models by a numerical PDE solver or by Monte Carlo simulation can be time consuming and requires expertise. The computational efficiency in the valuation of European options is of the highest importance when calibrating the model to option prices. A calibration techniques requires a very high number of option price evaluations with many different parameter settings. With an analytic expression pricing is time efficient.
As denoted in [44], 'the dynamics of the Heston model are not perfectly realistic, however, with an appropriate choice of parameters, all stochastic volatility models generate an implied volatility surface of roughly the same shape. Given the cheapness of the Heston model's computations, it is easy to see why the model is so popular.' Therefore we choose the Heston model in this research.

Heston Model

The dynamics of the Heston model can be described as follows,

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^1, \tag{2.2}$$

$$dv_t = \kappa(\theta - v_t)dt + \eta\sqrt{v_t} dW_t^2, \tag{2.3}$$

$$dW_t^1 dW_t^2 = \rho dt, \tag{2.4}$$

where:

- $\{S_t\}_{t\geq 0}$ is a continuous diffusion process of the stock price;
- $\{v_t\}_{t\geq 0}$ is the diffusion component of the return variance process;
- $\{W_t^1\}_{t\geq 0}, \{W_t^2\}_{t\geq 0}$ are two standard, correlated Brownian motion processes (with correlation parameter ρ) Heston [52].
- κ, θ,η, and ρ are parameters for the speed of mean-reversion, long-run mean, volatility of variance, and correlation between two standard Brownian motion $\{W_t^1\}_{t\geq 0}, \{W_t^2\}_{t\geq 0}$, respectively.

All the parameters, $\kappa, \theta, \eta, \rho$, are supposed to be time and state independent. Option prices are calculated by the Fourier transform pricing technique presented in Lemma A.1.1

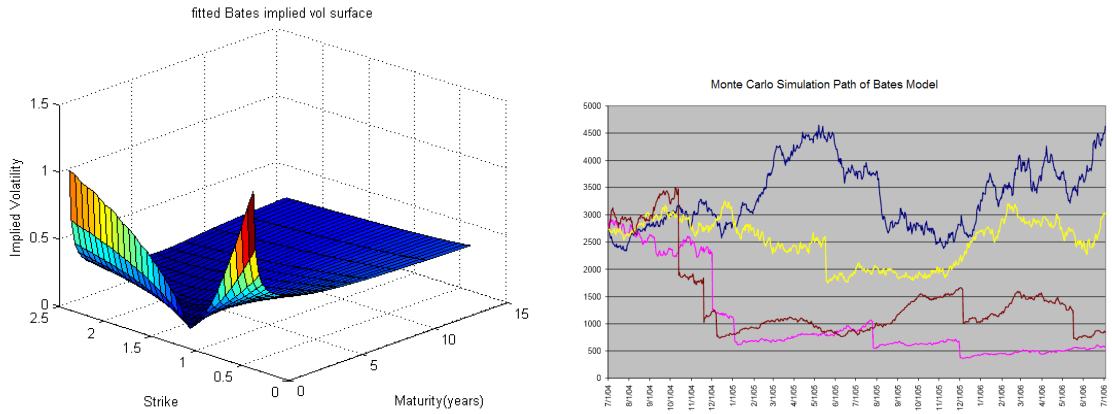

Figure 2.4: Left side: A schematic plot of the volatility surface in the Bates model. The skew is more pronounced at the short maturity end due to the inclusion of jumps. The wing-shaped surface is similar to Heston's, and the jumps do not affect the overall density of the price process significantly. The right-side figure shows some simulated paths generated by the Bates model with speed of mean reversion $\kappa = 4$, long run mean $\theta = 0.40$, vol-vol $\eta = 0.40$, correlation $\rho = -0.7$, initial variance $v_0 = 0.30$, jump intensity $\lambda = 1$, mean jump size $\mu = -0.50$, jump variance 0.10.

of Appendix A. Figure 2.3 presents an implied volatility surface generated by the Heston model.

Jump Extension of Heston's Model

The market implied volatility is typically more skewed at the extreme strike price of short maturity times than a Heston model can generate. This inability of the Heston model has motivated researchers to add a jump component to the stock price model, i.e., to create the correct asymptotic behavior of the implied volatility surface consistent to the market volatility surface. Bates [9] extended the Heston model by adding a simple log-normal distributed jump component to the Heston model:

$$\frac{dS_t}{S_t} = \mu dt + \sqrt{v_t} S dW_1 + j_{N_t} dN_t, \tag{2.5}$$

$$dv_t = \kappa(\theta - v_t)dt + \eta\sqrt{v_t} dW_t^2, \tag{2.6}$$

$$dW_t^1 dW_t^2 = \rho dt. \tag{2.7}$$

Here,

- j_{N_t} is the relative jump size of a Poisson jump process. It is log-normally, identically, and independently distributed over time, with unconditional mean μ_J and standard deviation η_J.
- N_t is a Poisson jump process with intensity λ, that is, $P\{dN_t = 1\} = \lambda dt$ and $P\{dN_t = 0\} = 1 - \lambda dt$.

Since the Poisson jump is independent with the Brownian motion, and the jump amplitude is continuous, the solution of this model can be obtained by adding the Fourier transformation of the jump part to the Heston model, such that $\phi(u) = \phi_H(u) \cdot \phi_J(u)$, where $\phi_J(u)$ is the Fourier transform of the Poisson jump component (Borak & Detlefsen [13]):

$$\phi_J(u) = e^{T[\lambda_j(e^{iu\mu_j - \frac{1}{2}\eta_j^2 u^2} - 1) - iu\lambda_j(e^{\mu_j + \eta_j^2/2} - 1)]} \tag{2.8}$$

Figure 2.4 shows an implied volatility surface generated by the Bates model.
One thing to remember is that the condition $2\kappa\theta \geq \eta^2$ should be satisfied any time. Otherwise, the path of the variance process v_t may become negative with a non-zero probability. This is undesirable and an incorrect feature which we have to avoid.

Chapter 3

Calibration Problems

The price to pay for more realistic models is the increased complexity of model calibration. Often, the estimation method become as crucial as the model itself

— Rama Cont, *Recovering volatility from option prices*[28]

One of the difficulties for applying stochastic volatility models in practice is the fact that there are unknown structural parameters which can not be implied from market data unambiguously. Until now, the way of obtaining these parameters is, still, immature in theory and practice. In this section, we therefore propose a numerically stable calibration procedure for stochastic volatility models, including the Heston and Bates model. The difficulties of the calibration procedure are discussed: Correlated parameters, multiple local minima, and noisy data. After the presentation of the calibration procedure, the quality of the fit of a single data set is shown. In order to improve this calibration locally (maturity-wise), time-dependent parameters are introduced. At the end of the chapter, we show how the model parameters change from day-to-day when calibrating the model to a time series of European option prices and present a fast intraday re-calibration procedure, which only costs 20 seconds.

3.1 Model Specification and Notations

Let's consider the Heston model, defined on a probability space $(\Omega, \mathcal{F}_T, P)$, with a two-dimensional random process $\vec{X}_t = (S_t, v_t)$, as the solution the following SDE,

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^1 \\ dv_t &= \kappa(\theta - v_t)dt + \eta\sqrt{v_t} d(\rho W_t^1 + \sqrt{1-\rho^2} Z_t) \\ dW_t^1 dZ_t &= 0 \end{aligned}$$

where

- $\{S_t\}_{t\geq 0}$ is the continuous diffusion process of the stock price;
- $\{v_t\}_{t\geq 0}$ is the diffusion component of the return variance process;
- $\{W_t^1\}_{t\geq 0}, \{W_t^2\}_{t\geq 0}$ are two standard, but correlated, Brownian motions (with correlation parameter ρ) [52].
- κ, θ,η, and ρ are parameters denoting the speed of mean-reversion, long-run mean, volatility of variance, and correlation between two standard Brownian motions $\{W_t^1\}_{t\geq 0}, \{W_t^2\}_{t\geq 0}$, respectively.

The parameter set is defined as follows:

$$\xi = (v_0; \chi) = (v_0; \kappa, \theta, \rho, \eta)$$

This division is not superficial, since parameter v_0 appear to be a 'hidden' state variable and the χ is a vector of structural parameters. The reason for this definition of the parameter set is that the stability of the calibration procedure is improved by optimizing over a lower-dimensional subspace. The market price of ith benchmark option are denoted by C_i^{mkt}. $C_i(\xi_j)$ indicates the model generated option price under model parameters ξ_j at the jth Monte Carlo iteration ($j \in [1, \nu]$) .

3.2 Description of Market Data

The data considered in this research is the implied volatility data of daily Eurostoxx50 index options. The sample period extends from February 1st, 2007 until September 17th,

2007[1]. The discount rate is obtained by averaging the risk neutral forward rate in the same period.

3.2.1 Simple Test for the Noise in Data

Because the prices are set by the traders who have correctly or incorrectly used financial models, market data may contain 'noise', e.g. errors or misleading information. Possible reasons are:

1. The bid and ask spread,
2. Illiquidity of the options (with far in- or out-of-the-money strikes),
3. Errors in recording.

We have developed a simple test to identify the quality of data. The testing technique is based on the concept of a *butterfly spread* option. The *butterfly spread* option maturing at time T represents a portfolio of three European calls on consecutive strikes,

$$B_0(S_0, K, T, \epsilon) = \frac{1}{\epsilon^2}(C_0^{mkt}(S_0, K-\epsilon, T) + C_0^{mkt}(S_0, K+\epsilon, T) - 2C_0^{mkt}(S_0, K, T)), \quad (3.1)$$

for some $\epsilon > 0$. Here $C_0(S_0, K, T)$ represents the market price of a European call with the current stock price S_0, strike K, and time to maturity T. So, we buy a call struck at $K+\epsilon$, buy another call at $K-\epsilon$, and sell two call options at K, with all calls maturing at the same time. When $T \to 0$, the payoff function for the butterfly spread converges to a Dirac-delta function in the limit $\epsilon \to 0$ [1] (see Figure 3.1):

$$\begin{aligned}\delta_\epsilon(S_T - K) &= \lim_{\epsilon\to 0} B_0(S_0, K, T, \epsilon) \\ &= \lim_{\epsilon\to 0} \frac{1}{\epsilon^2}(C_0^{mkt}(S_0, K-\epsilon, T) + C_0^{mkt}(S_0, K+\epsilon, T) - 2C_0^{mkt}(S_0, K, T)) \quad (3.2)\end{aligned}$$

For a Dirac-delta function, the following property holds,

$$\lim_{\epsilon\to 0}\int_0^\infty \delta_\epsilon(S_T - K)f(K)dK = \int_0^\infty \delta(S_T - K)f(K)dK = f(S_T), \quad (3.3)$$

[1]We thank Rob and Frank at Rabobank International, London office for their help in getting this data set

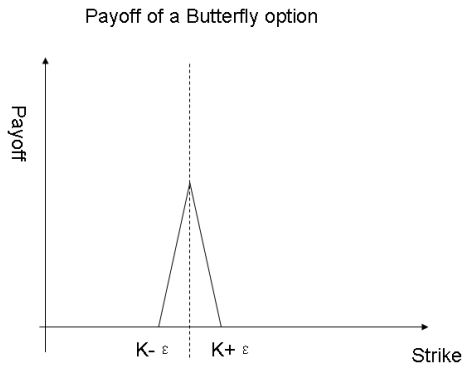

Figure 3.1: The shape of the payoff function of a butterfly option resembles the dirac-delta function.

for any $S_T > 0$ and any continuous function $f(x)$. From the risk-neutral pricing formula we have that

$$\lim_{\epsilon \to 0} B_0(S_0, K, T, \epsilon) = e^{-rT} E_0^Q[\delta_\epsilon(S_T - K)].$$

In particular, we find that in the limit $\epsilon \to 0$,

$$\begin{aligned}
\lim_{\epsilon \to 0} B_0(S_0, K, T, \epsilon) &= \lim_{\epsilon \to 0} e^{-rT} E_0^Q[\delta_\epsilon(S_T - K)] \\
&= e^{-rT} \lim_{\epsilon \to 0} \int_0^\infty p(S_0, 0; S_T, T) \delta_\epsilon(S_T - K) dS_T \\
&= e^{-rT} \int_0^\infty p(S_0, 0; S_T, T) \delta(S_T - K) dS_T \\
&= e^{-rT} p(S_0, 0; K, T) > 0 \qquad (3.4)
\end{aligned}$$

where $p(S_0, 0; K, T)$ is the risk-neutral density that the stock price S_T equals K at time $t = T$.

From inequality (3.4) it can be seen that if we construct a butterfly option by 3 consecutive call options we should get an approximation of the risk-neutral probability, which is non-negative. If the price of the butterfly option is negative, the associated data point (three option prices) contains noise. By constructing the butterfly option from the market prices, we get the plot in Figure 3.2. This plot suggests that options with maturities over 3 years tend to violate the arbitrage-free bound drastically. Therefore we suggest to use only option prices with shorter than 3 years maturities.

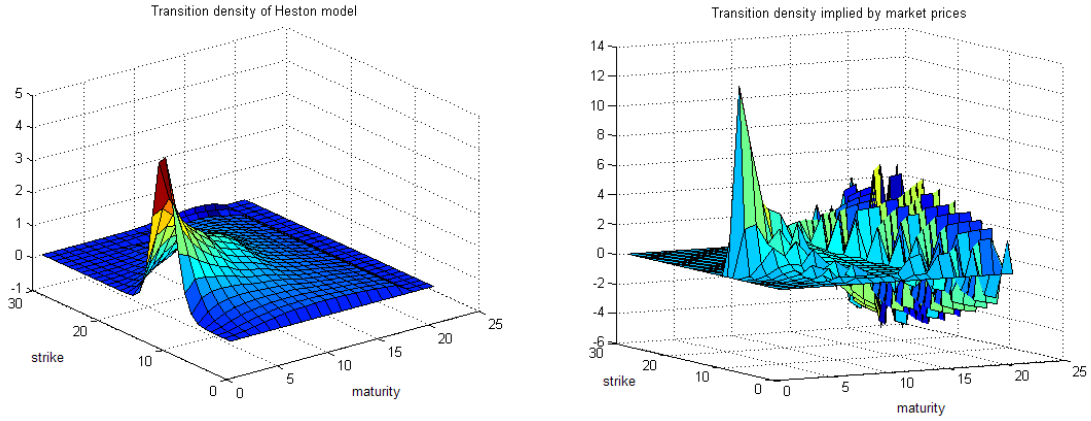

Figure 3.2: The smooth transition density surface generated by a continuous diffusion model (Heston), in contrast to the transition density generated by the noisy real market data.

3.3 An Inverse Problem Formulation for Calibration

The 'market implied' approach for model calibration procedure relies on the existence of analytical or semi-analytical expression for the prices of benchmark instruments. The unknown parameters of the financial models are found by inverting the pricing formulas that are the closest to the market prices. This method is called *inverse problem formulation*.

3.3.1 A Least Square Problem Formulation

In the sense of least-square error, we minimize the errors between prices predicted by the model and the market option data $\hat{\epsilon}_i(K_i, T_i; \xi) = |C_i^{mkt} - C_i(\xi_j)|$. We define the objective function to be a scalar-valued function,

$$G(\xi) = \sum_{i=1}^{N} l(\hat{\epsilon}_i(K_i, T_i; \xi)),$$

where l is a suitably chosen norm. The optimal parameter set, $\hat{\xi}$, is the one with minimal $G(\xi)$, i.e.,

$$\hat{\xi} = arg \min_{\xi \in \Omega} G(\xi).$$

In practice, a numerical optimization procedure is required to find the global optimum. Usually, the norm l is chosen to be a 'quadratic' norm (ℓ_2-norm). The main criticism

when choosing a quadratic norms is that it implicitly assigns more weight to the 'expensive' options (e.g., ITM options and long-term options), and less weight to the short-term or OTM options. This phenomenon can be eliminated by the *relative (percentual) pricing error*,

$$l(\hat{\epsilon}_i(K_i, T_i; \xi)) = \frac{|C_i(\xi_j) - C_i^{mkt}|}{C_i^{mkt}}$$

However, by doing so, we would favor cheaper options, especially short-term, extremely skewed options[6]. This may lead to a somewhat better fit of short-term options at the cost of the fit of ITM and long-term options.

It is known [44] that the Heston model only has a limited ability of generating extreme skews for short maturities, therefore the model error related to this will be large. The choice for the relative percentual pricing error tends to give a large overall model error. As we can see from Table 3.1, the relative pricing error seems to overestimate the correlation parameter, and to underestimate the vol-vol parameter, η, in order to fit the extremal one-sided skew at the short maturity end. Based on this observation, we stay with the conventional *quadratic error norm* as the norm in the objective function.

In detail, the optimization procedure is described as follows:

$$\min_{\chi} G(v_0; \chi) = \min_{\chi} \sqrt{\sum_{i=1}^{N} \omega_i \{C_i(\xi_j) - C_i^{mkt}\}^2} \tag{3.5}$$

where $G(v_0; \chi)$ is the objective function, χ is a 4-dimensional vector of structural parameters, $C_i(\xi_j)$ and $C_i^M(K_i, T_i)$ represent the i^{th} option prices from the model and the market, respectively, with strike K_i and maturity T_i. N is the number of options used for calibration. Weight ω_i is used during the optimization procedure. Cont [28] suggests to use the BS-Vega as the weight. In the appendix D, we provide a theoretical proof of such proposal. However, we also observe that this choice of weights favors ATM options while it is insensitive to the extreme skew at the far OTM and ITM strikes (vega peaks around ATM strikes). Therefore, we choose uniform weights $\omega_0 = \omega_1 = \cdots = \omega_N$ such that the weights rescale the total square error to 50, in which region the global optimization

algorithm (to be introduced in the next chapter) works most effectively,

$$\omega \sum_{i=1}^{N} \left\{ C_i^{mkt} - 0 \right\}^2 = 50$$

Table 3.1: Different norms give rise to different optimal parameters after fitting the Heston model to Eurostoxx50 call options for May 30th, 2007

Objective function	**Maturities up to**	**Solutions**$(\kappa, \theta, \eta, \rho, v_0)$
Relative Err.	14 years	$(1.145, 0.011, 0.186, -0.977, 0.044)$
Quadratic Err.	14 years	$(0.018, 0.559, 0.203, -0.657, 0.030)$
Relative Err.	3 years	$(3.302, -0.004, 0.056, -1, 0.038)$
Quadratic Err.	3 years	$(0.858, 0.050, 0.418, -0.654, 0.035)$

3.3.2 A Well-Posed Formulation and Regularization by Relative Entropy

The resulting problem formulation from previous section is ill-posed, since it is analogous to the problem of determine the coefficients of an unknown function from discretely observed noisy inputs. The solution for this problem, may be unique, or may not be unique, or does not have to have solution at all. Therefore, regularization (smoothing) is needed.

The standard way of doing this is by adding a penalty function to the square error term, so that we are minimizing over the regularized function

$$J_\alpha = G + \alpha S$$

where G measures the accuracy and S measures the smoothness. The parameter α is used to adjust the trade-off between accuracy and smoothness. Here we reformulate the problem in a well-posed fashion by combining the relative entropy I.

As is analyzed in the Cont [27], the relative entropy method is used for solving ill-posed inverse problems in financial model calibration. In this application, a not sufficiently specified problem is solved by using relative entropy to fill in the missing information. The relative entropy measure the 'information distance' of the calibrated model to the prior distribution. In the context of financial model calibration, Buchen and Kelly [16]

suggest to choose the following functional form for the relative entropy

$$I(p|p_0) = \int p(\xi) \log(\frac{p(\xi)}{p_0(\xi)} d\xi$$

where p_0 is the prior distribution of a set of parameters ξ.
In the case of stochastic optimization, the uniform prior $p_0 = \frac{1}{\nu}$ is used to indicate our ignorance of the true solution

$$I(p|p_0) = \sum_{j=1}^{\nu} p_j(\log p_j - \log \frac{1}{\nu}) = \log \nu + \sum_{j=1}^{\nu} p_j \log p_j$$

Recall that the function $f(x) = x \ln x$ is strictly convex in the domain of $x \in R^+$, so that it is natural to use the relative entropy as the penalty function to regularize the objective function. The use of relative entropy for calibration has been justified by solid economic arguments. Samperi [84] shows that the relative entropy penalization for model calibration is mathematically equivalent to using a penalization function in the probability space and the calculation of Arrow-Debreu prices by utility maximization. See Samperi [84] and Avellaneda [5] for detailed and in-depth discussions.

3.3.3 Regularized Objective Function and Convex Optimization

Following Avellaneda's [5] argument, we reformulate the problem to be a mini-max programming problem (the dual formulation of the least square problem) by introducing Lagrange multipliers $(\lambda_1, \ldots, \lambda_N)$

$$\min_{\lambda} \Big[\max_{p} \Big(- I(p|p_0) + \sum_{i=1}^{N} \lambda_i \big(\sum_{j=1}^{\nu} p_j C_i(\xi_j) \Big) - \sum_{i=1}^{N} \lambda_i C_i^{mkt} \Big] \tag{3.6}$$

where N denotes the number of option prices obtained from the market.
Firstly, let us fix λ and seek the density p that maximizes this 'augmented Lagrangian'. The calculation of the first-order optimality condition by Cover & Thomas [32] shows that for each λ, the elements of the optimal probability vector p are of the Boltzmann-Gibbs form

$$p_j = p(\omega^{(j)}) = \frac{1}{Z(\lambda)} \exp \Big(\sum_{i=1}^{N} \lambda_i C_i(\xi_j) \Big). \tag{3.7}$$

where the $Z(\lambda)$ is the normalization factor

$$Z(\lambda) = \frac{1}{\nu}\sum_{j=1}^{\nu}\exp\left(\sum_{i=1}^{N} C_i(\xi_j)\lambda_i\right)$$

Secondly, substitute Equation 3.7 into Equation 3.6, we have

$$\begin{aligned} V(\lambda) &= -I(p|p_0) + \sum_{i=1}^{N}\lambda_i\big(\sum_{j=1}^{\nu} p_j C_i(\xi_j) - C_i^{mkt}\big) \\ &= \log(Z(\lambda)) - \sum_{i=1}^{N}\lambda_i C_i^{mkt} \\ &= \log\left\{\frac{1}{\nu}\sum_{j=1}^{\nu}\exp\left(\sum_{i=1}^{N} C_i(\xi_j)\lambda_i\right)\right\} - \sum_{i=1}^{N}\lambda_i C_i^{mkt} \end{aligned}$$

So that we can determine the optimal Lagrange multipliers λ^* by minimizing over the function $V(\lambda)$. Given the technical assumptions proved by Samperi [84] that $V(\lambda)$ is twice continuous differentiable and strictly convex over λ, we know that the first order conditions for its unique minimum for λ_k are

$$\begin{aligned} \frac{\partial V}{\partial \lambda_k} &= \frac{1}{Z}\frac{\partial Z}{\partial \lambda_k} - C_k^{mkt} \\ &= \frac{1}{Z}\frac{1}{\nu}\sum_{j=1}^{\nu} C_k(\xi_j)\exp\left(\sum_{i=1}^{N} C_i(\xi_j)\lambda_i\right) - C_k^{mkt} \\ &= \mathbb{E}\{C_k(\xi)\} - C_k^{mkt} = 0 \end{aligned}$$

The expectation sign in the last identity means the expected (average) model generated price for kth benchmark instrument (vanilla option in this case) over all candidate parameter sets ξ (selected by the Monte Carlo procedure).

The above derivation shows that the calibration problem can be transformed into the problem of finding the minimum of a strictly convex function $V(\lambda)$ of N-vector of Lagrange multiplier, which problem can be solved easily by standard gradient method. This dual objective function can also be solved efficiently by Sequential Quadratic Programming (with L-BFGS [100] update) implemented in nag_opt_nlin_lsq optimization subroutine.

Algorithm 1 (A Calibration Procedure for Lagrange Dual Objective Function). *1. Construct a Monte Carlo to sample the parameter space.*

2. *Compute the model generated option price matrix* $\{C_{ij} = C_i(\xi_j), i = 1, \ldots, \nu, j = 1, 2, \ldots, N, \}$.

3. *Using Sequential Quadratic Programming optimization routine to minimize the function* $V(\lambda)$.

4. *Compute the posterior distribution of model parameters* $p_j, j = 1, 2, \ldots, \nu$ *with the optimal* λ.

3.4 Numerical Methods for Option Valuation

We did not yet describe how to generate the model prices efficiently and accurately. This will be the topic of this section. For a complex model, like the Heston model, closed form solutions do not exist, so we have to approximate its solution a numerical technique, like,

- A finite difference solution of the corresponding partial differential equation (PDE),
- Monte Carlo simulation method combined with a variance reduction technique,
- Fourier transform technique.

There is a long list of the pros and cons of each of these methods, which we do not intend to discuss in details. We just highlight some important aspects here. The PDE method is a flexible method which can be used for many payoffs, e.g. European options or certain path dependent derivatives. The drawback is that we have to approximate the option prices on a grid. Accurate pricing requires a substantial amount of grid points, which leads to a considerable amount of extra work Lindström & Ströjby [66]. The PDE method is somewhat expensive.
The Monte Carlo method is the most general of the three techniques. However, its convergence is very sensitive to simulation noise. Therefore it also requires long computation times.
In this project, therefore the Fourier transformation technique has been used to evaluate the model option prices. This method is both fast and accurate. Its major technical difficulty lies in the derivation of a characteristic function, i.e., the Fourier transform of

the risk-neutral density function. See Carr and Madan [23] for further details. The Fourier transformation technique can take advantage of a very fast numerical algorithm called the *Fast Fourier Transformation* (FFT) technique, which drastically improves the numerical efficiency of the calibration procedure, as the pricing technique is called for many times during the calibration. The implementated FFT method is described in some detail in Appendix A.

3.5 Parameter Estimation for a Single Dataset

Our first calibration objective is the best possible fit (the smallest square error) for a single option data set. In order to enhance the numerical efficiency and stability of the calibration procedure, it is separated into different steps, described in detail in the following subsections.
Firstly, we estimate the initial variance, v_0, by using ATM implied volatilities. Secondly, we apply a global optimization procedure to extract optimally fitted time-independent parameters, χ. At this stage, we already achieve a good fit. Thirdly, we add jumps to further improve the fit for the time-independent case.

Step 1: Hidden State Variable Estimation

It is desirable to estimate the hidden state variable, v_0, first. The estimation of v_0 can be performed with the ATM implied variance, based on the following lemma from Gatheral [44]:

LEMMA 3.5.1 (Term structure of the Black-Scholes implied volatility in the Heston Model)**.**

$$\sigma^2_{ATM}(T) \approx \frac{1}{T}\int_0^T \left[(v-\theta')e^{-\kappa' t} + \theta'\right]dt \tag{3.8}$$

$$= (v-\theta')\frac{1-e^{-\kappa' T}}{\kappa' T} + \theta' \tag{3.9}$$

where $\kappa' = \kappa - \frac{\rho\eta}{2}, \theta' = \frac{\theta\kappa}{\kappa'}$.
The ATM BS implied variance $\lim_{T\to 0}\sigma^2_{BS}(T)|_{ATM} \to v_0$.

The proof can be found in Appendix B.

Table 3.2: Many competing minima after fitting the Heston model to the call option of Eurostoxx50 on May 30th, 2007

Solutions$(\kappa, \theta, \eta, \rho, v_0)$	**SQ Difference**	**Regime**
$(1.3678, 0.08, 1.3102, -0.6708, 0.0257)$	$0.0173(/51.5632*)$	High vol high mean reverting
$(0.0385, 0.3074, 0.2177, -0.6549, 0.0296)$	$0.0026(/51.5632)$	Low vol low mean reverting
$(0.0481, 0.2556, 0.218, -0.6631, 0.0295)$	$0.0026(/51.5632)$	Low vol low mean reverting
$(0.0425, 0.2822, 0.2172, -0.6621, 0.0296)$	$0.0026(/51.5632)$	Low vol low mean reverting
$(1.4829, 0.0872, 1.7048, -0.6288, 0.0104)$	$0.0175(/51.5632)$	High vol high mean reverting

* the summation of all market option prices across all maturity and strikes.

Practically this means that if we assume that the stock process follows the Heston dynamics, then v_0 should be consistent with the short dated ATM volatility. We assume from the lemma that there is a linear relationship between the initial variance, v_0, and the BS implied variance returned by the Heston model. Experimentally we found that for a very short maturity, say 1 week (0.02 years), the relationship is not really linear, as then the implied variance tends to zero as the price implied by the Heston model converges to the payoff function. In the range of 1 to 2 months maturities, this relationship is close to linear as can be seen from Figure 3.3. From the results in this figure, we conclude that *the Black-Scholes implied variance of ATM options of about 2 months maturity is a satisfactory estimate for the initial variance, v_0, in the Heston model.*

Step 2: Stationary Parameters and Global Optimization

Besides the noisy input data, issues also complicate the optimization problem, are

- Multiple local minima, i.e., several minima that are separated by a ridge,
- Expensive function evaluations.

1. Multiple local minima

Because of the existence of model error and noise in the data, the landscape of the objective function is rugged, it is neither convex nor of any regular shape. Many local minima exist possibly separated by barriers. Table 3.2, for example, lists some of the local minima found during the optimization of the Heston model. Here we can see that if the volatility of the volatility parameter is high and the speed-of-mean-reversion is also high, then the consequence is that the dynamics of the stock price displays a sudden 'burst of

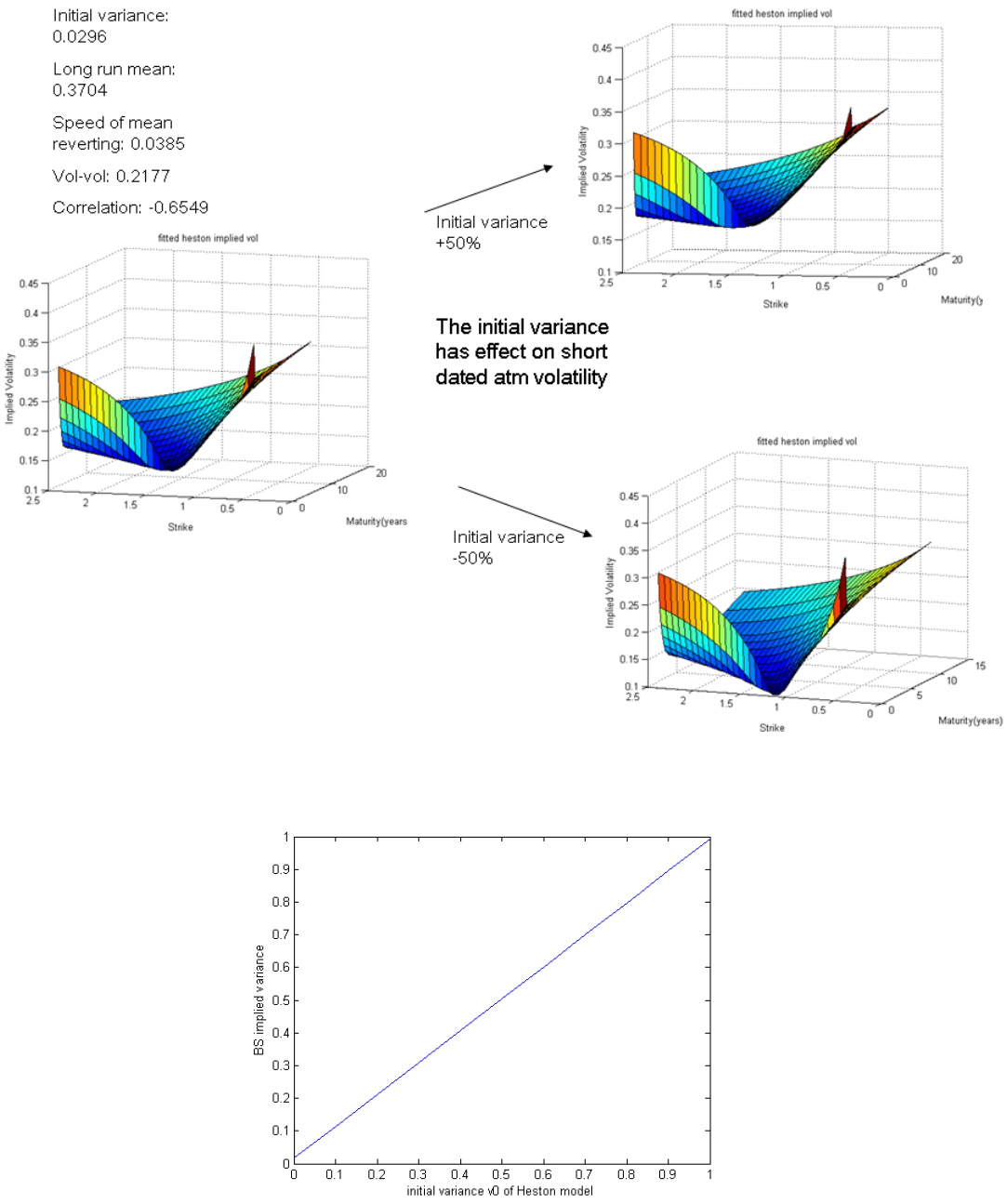

Figure 3.3: The top figure shows the effect of shifting the initial variance. The far ITM and OTM implied volatilities are not affected by changing v_0 neither is the long dated ATM vol. The bottom figure shows the relationship between the ATM implied variance of maturity T=0.14 (between 1 and 2 months), and v_0, which is more-or-less linear to the initial variance v_0.

volatility' which quickly reverts to its usual level. From the calibration point of view, there may be a set of parameters in the high vol-vol, high speed of mean reverting regime which gives a compatible fit to a set of parameters in the lower vol-vol, lower mean reverting regime. This feature is sometimes characterized by the phenomenon that 'parameters that are highly correlated'. In the next chapter, we develop a *partial resampling scheme* which groups the highly correlated parameters together in order to improve the convergence to a global optimum.

Due to the noise which is still present in the option price data, the objective function is subject to high-frequency, low amplitude noise.

The issues above prevent us from using simple finite differences-based gradient algorithms, like the quasi-Newton algorithm, as they converge to a local optimum Booker & Dennis, 1998 [12].

2. Expensive function evaluation

Although the option pricing method is efficient (based on FFT algorithm), in order to calculate the square error for one parameter set, a full option price matrix with 20 maturities and 32 strikes and thus 640 option prices, has to be calculated. The computational time for the square error of one parameter set is approx. 0.8 seconds, thus 2000 iterations take almost half an hour. Unfortunately, on the trading floor, people cannot wait this long for an answer. So the major challenge here is to develop an optimizer which finds the global optimum in a high dimensional space in less than 2000 function evaluations.

These difficulties render most of the common optimization algorithms inefficient. It motivates us to build an more efficient global optimization algorithm. The details are given in the next chapter.

Result of Step 2: The Quality of In-Sample Fit

Fitting the Heston prices to the European option prices matrix of maturities longer than 2 weeks (32 strikes $\times$ 18 maturities) yields a square error of 0.0026 (see Table 3.2) out of 51.5632 in total (which is the summation of all market option prices across all maturities and strikes). The ATM fit and the total square difference surface are shown in Figure 3.4. As we can see from that figure the ATM volatility fit is highly satisfactory for long maturities – the difference between the Heston ATM volatility and the market implied volatility for maturities longer than two years is less than 1%. However, as mentioned

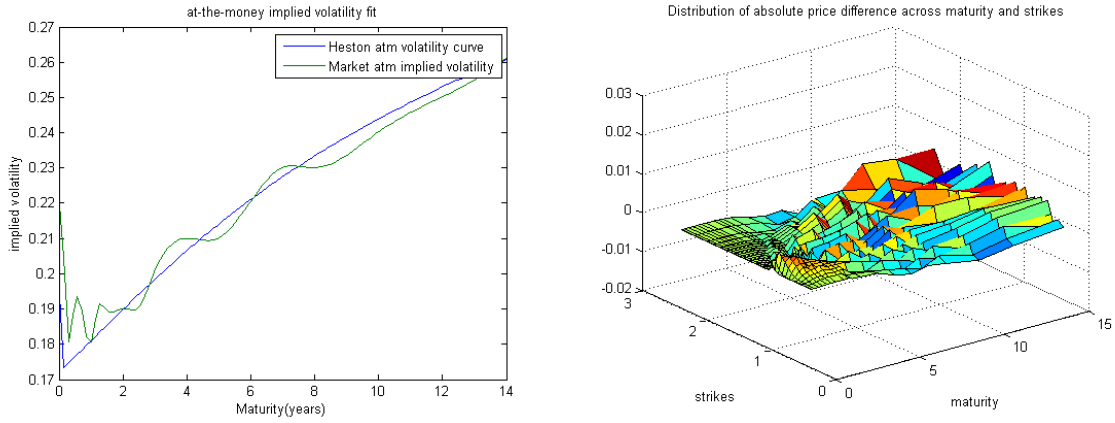

Figure 3.4: The quality of static fit of the global optimal parameter set (0.0385, 0.3074, 0.2177, -0.6549, 0.0296). The left-side figure shows the surface of square differences of the model prices with respect to market prices. The right-side figure displays the distribution of price differences between model and market prices with the parameters above.

earlier, the difference at the short end is not so satisfactory. We take a closer look at the fit to the volatility skew for each maturity, presented in Figure 3.5. Here, we see that the fit is generally good for longer maturities, and of less quality for short maturities. Moreover, the far OTM volatility skew seems to fit worse than the far ITM skew. The OTM options are cheaper, thus have less weight, than the ITM options. A conclusion here is that after the global calibration procedure we obtain a set of time-independent parameters, with the following properties

1. a reasonably good fit to the ATM implied volatilities, which contains the most important pieces of information.

2. it produces the qualitative correct shape for the ITM and OTM implied volatility skews.

Step 3: Adding Jumps

As we can see from Figure 3.5, the fit of the Heston model to the short end of the implied volatility surface is not satisfactory. Therefore, we can add jumps to the Heston model (leading to the Bates model) to further improve the fit to the market prices. Figure 3.6 presents the improvement of the fit to the short end implied volatility surface. At the same time, the fit to the long term implied volatility is not worse than that of the Heston

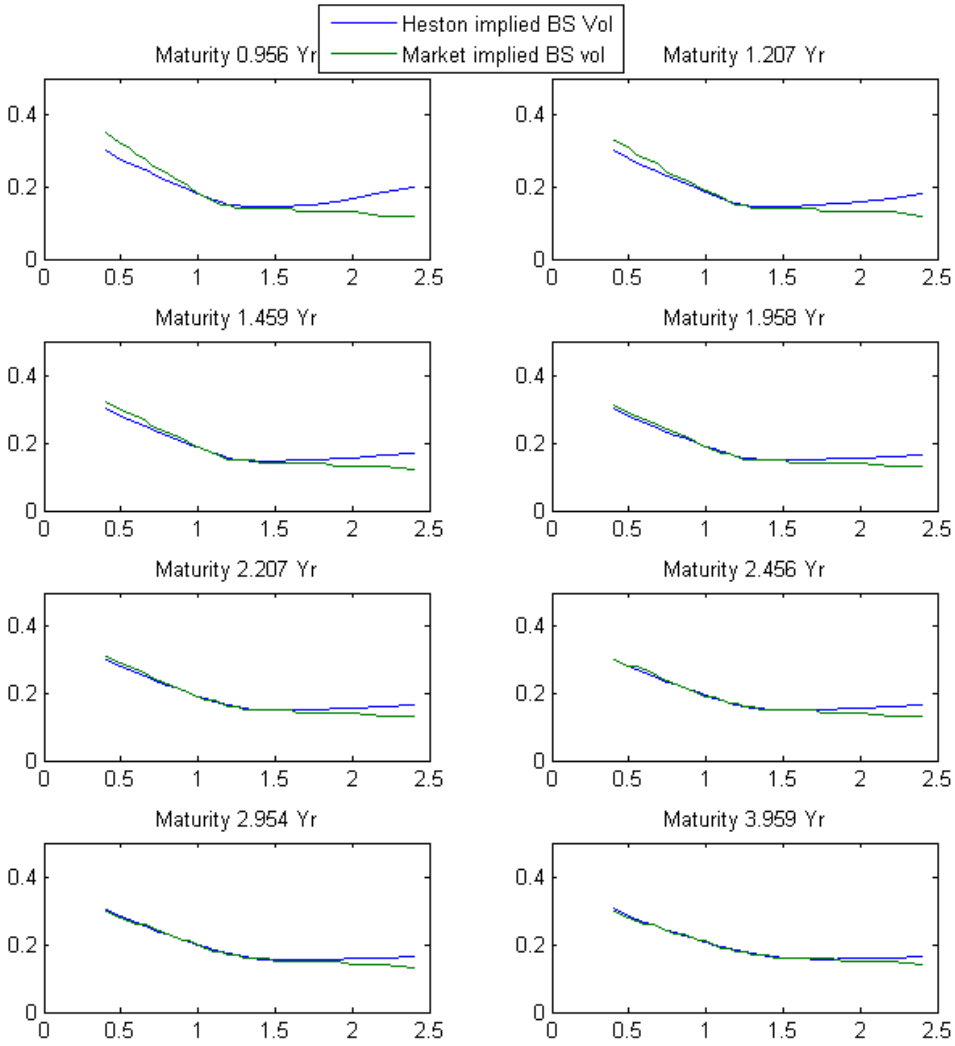

Figure 3.5: Plot of implied volatility skews across all listed maturities produced by time-independent parameters. It doesnot yield a satisfactory fit for the short maturities. However, the fit gets better as the maturity moves further in time.

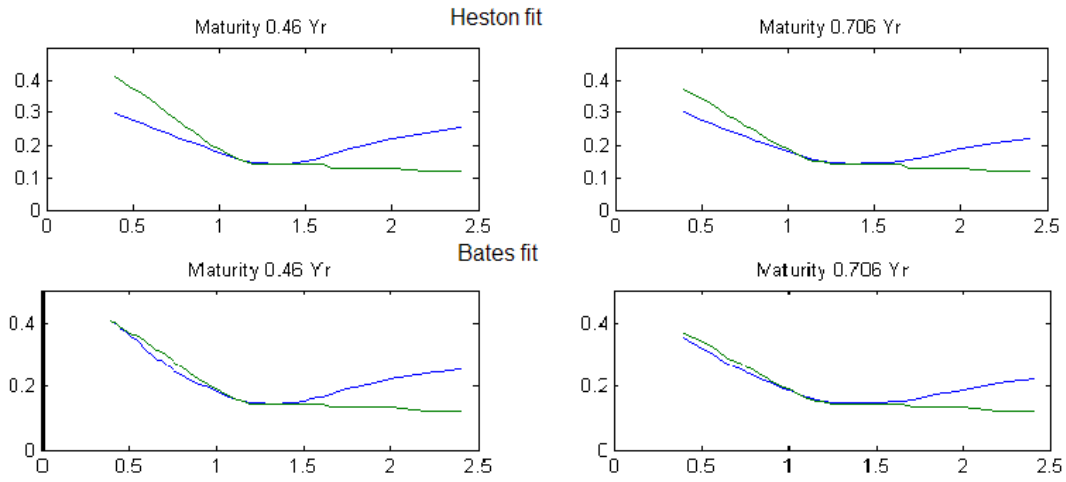

Figure 3.6: After adding jumps (Bates model) the fit to the short end implied volatility surface is significantly improved.

model. We notice that adding jumps is interesting as long as the optimization algorithm is capable of handling this larger, 8-dimensional, parameter space.

Time-Dependent Parameters and Improved In-Sample Fit

We have seen that the Heston model captures the main features of the market implied volatility surface. However, the pricing errors are distributed across several maturities and strikes. An idea for improvement is to make the parameters *time-dependent* in order to improve the fit. However, a naive implementation would result in highly fluctuating parameters, indicating that an unnatural evolution of stock prices is assumed. We found that only the correlation, ρ and the vol-of-vol parameter η need to be made time-dependent. The indicating result for calibration is a significant further reduction of the pricing error, although we keep the time dependency of the parameters to a minimum. As the initial variance, v_0, mainly matters for the short end of the volatility surface, there is no need to making v_0 time dependent. By imposing a consistency condition on the forward variance curve, Buehler [19] showed that the mean reverting rate κ cannot be

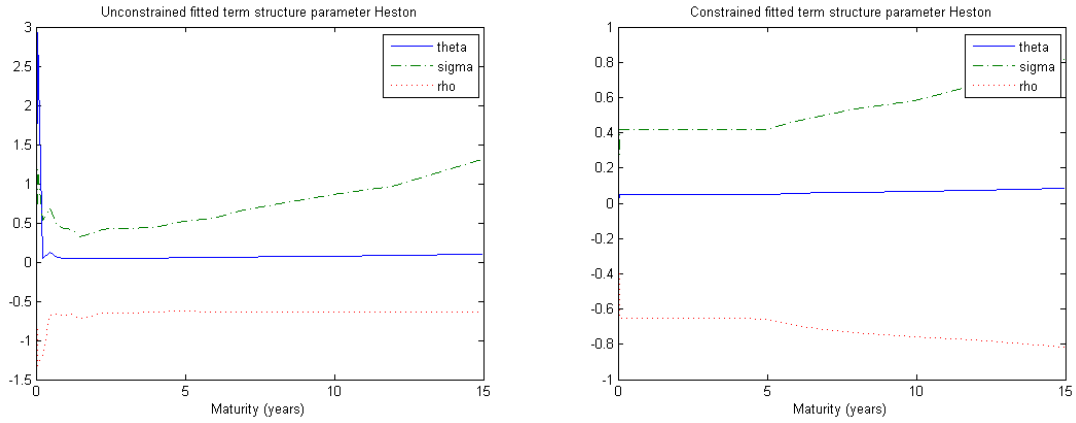

Figure 3.7: The time-dependence of parameters in one dataset. Left-side figure shows the unconstrained fit; The right-side the constrained fit for the time-dependent parameters.

time-dependent. Otherwise, the forward variance, which can be replicated by a strip of put and call options, implied by the model would not be consistent with the market prices of the products called variance swaps. Fixing κ increases the stability of the parameters through all maturity slices.

With time-dependent parameters θ, η, ρ and unconstrained optimization, we plot the resulting three, unsteady, parameters for one Eurostoxx50 implied volatility surface on May 1st, 2006 in Figure 3.7. We observe unnatural parameters at the short end of maturity due to the strong negative skew there. For longer maturities, we see a more stable ρ and long run mean θ, whereas only the vol-vol parameter is varying significantly in time. In order to remove the undesired parameters at the short end, we then constrain the feasible parameter region to $[0.001, 0.1]$ for θ, $[0.01, 1]$ for η and $[-1, 0]$ for ρ and performed *a constrained optimization* in the second experiment. Here, we see that the ρ also increases to compensate for the constrained η when producing the skew (the skew is linear to the product of ρ and η).

From these two pictures we conclude that a time dependent vol-vol parameter, η, and correlation, ρ, are important to improve the fit. For some derivatives that indicate exposure to correlation or vol-vol risk, the time dependency of η, ρ, or both, is important, as we can hedge against the parameter risk which arises from this time-dependent movement.

3.6 Intraday Parameter Estimation

After we obtained the model parameters by calibration we can apply a 'model hedge' to hedge the risk of derivatives we sold, however, there is still the risk of unstable model parameters (see Figure 3.8). If the option prices and hedge parameters implied by the daily recalibrated Heston model differ significantly from a previous day, the model needs to be re-calibrated and the hedge has to be rebalanced.

There are, in general, two solutions for dealing with the risk of changing parameters. One is the so-called 'parameter hedging' as described on pp. 74 of Buehler's PhD thesis [20]. The idea is to calculate the derivative of the price with respect to the parameters and then neutralize the derivatives by hedging with liquid vanilla options. It is important to note that such a trading strategy is not self-financing anymore. Which option to choose and which position to take in the options is determined by a constrained optimization procedure. The procedure of parameter hedging is to optimize in the space of options (each parameter sensitivity is compensated with an option) such that the overall portfolio sensitivity is the smallest. The optimality constraints are usually total cost (including the transaction cost) and liquidity. Additional constraints can be the size of the position in these options.

A fast and robust solution is to re-calibrate the model everyday and to rebalance the hedge accordingly. Here, the speed of the calibration is crucial. Our calibration procedure (without time dependence) takes about 10 minutes (calibrate to the option prices of 32 strikes and maturities up to 3 years), which is certainly too long in a real-time environment. Therefore, we here also develop a fast re-calibration procedure. The idea here is to fix the speed of mean reversion, κ, to be the same as yesterday's and to determine the initial variance, v_0, from the 2 month ATM implied volatility. By doing so, the complexity of the optimization problem is greatly reduced and the optimization algorithm finds a global optimum in merely 20 seconds. The price to pay is a decrease of the quality of fit (the measured square error increased from 0.2% to 0.3%) compared to the full calibration procedure.

Buehler [21] showed that choosing different values for the speed of mean reversion, κ, does not reduce the hedging performance significantly. In fact, the hedging performance under $\kappa = 0.01$ and $\kappa = 5$ are almost the same in [21]. Hence, this re-calibration procedure is fast without sacrificing too much parameter quality (since the in-sample fit and hedging

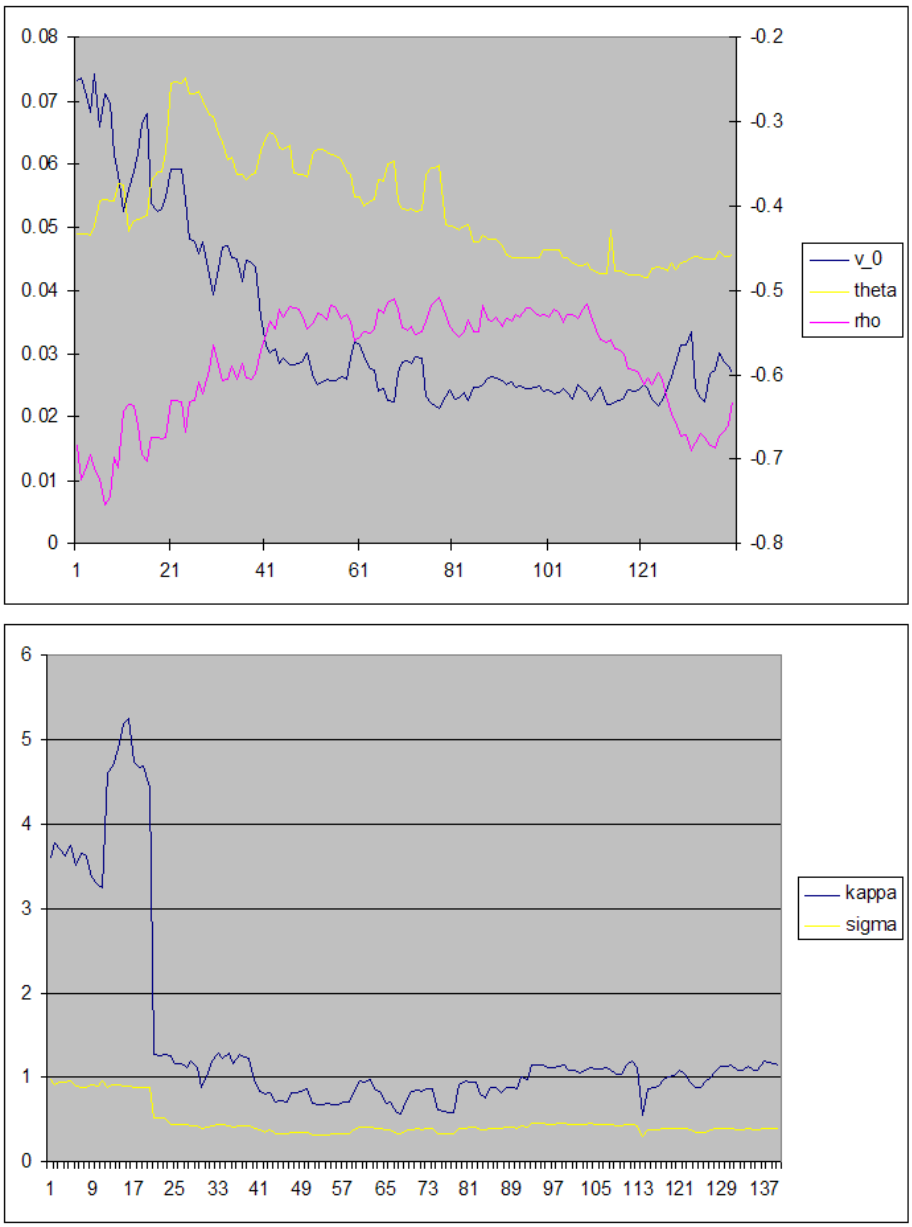

Figure 3.8: The plot of time-varying Heston parameters obtained by calibrating to a time series of cross-sectional European option prices from Feb 1st 2007 - June 1st 2007. The speed of mean reverting κ and correlation ρ are more volatile than the other parameters.

performance are still satisfactory).

3.7 Conclusion

In this chapter we described the calibration procedure for the Heston model. We first estimate the initial variance, v_0, from the ATM implied volatility, then perform a global optimization procedure to obtain the best fitted time-independent parameters. After calibration, we found that the model fits the long maturity implied volatility well. However, it is difficult to generate a sufficiently steep skew for short maturities within the Heston model. This can be overcome by adding jumps to the model or by introducing some form of time dependence in the parameters (only θ, η, ρ may vary in time). The latter one has the unpleasant consequence of time-dependent smile surfaces. For products with a high parameter sensitivity, it is necessary to make the parameters time-dependent so that the risk of changing parameters can be controlled. To deal with the risk of parameters changing from day-to-day, we suggest to recalibrate the model in a daily basis to control the risk of changing parameters from day to day. A fast re-calibration can be obtained by fixing the speed of mean reversion, κ and to read the initial variance, v_0, from ATM implied volatility.

Chapter 4

Hybrid Stochastic Approximation Search Algorithm

> *. . . relatively simple algorithms may perform as well as – or better than – other "exotic" algorithms that may have received a disproportionate amount of attention. A correct choice of algorithm may lead to significant cost savings through the production of a better solution to the problem and through savings in the effort required for implementation.*
>
> — James C. Spall *Introduction to Stochastic Search and Optimization*[93]

This chapter starts with a general discussion of difficulties in, so-called hard optimization problems in general and the concept of 'state space of complex systems'. The dynamics within the complex system is brought down to the theoretical ground of Markov proces, since we found that almost all the known stochastic optimization processes, e.g. genetic algorithm, simulated annealing, can be understood as a special case of Markov process. Here in this chapter we describe two Markov processes with distinct dynamics in the state space: First one, Wang-Landau algorithm, is capable of escaping from the local entrapment; while the second one, partial resampling, reduces the big fluctuation generated by a very general movement. They are combined in a multi-level structure so that the level 1, WL algorithm, guarantees that the dynamics of the algorithm is global and the level 2, partial resampling, increases the efficiency of the stochastic search by steering the algorithm with problem specific information. In the end of this chapter implementation issues are discussed.

4.1 No Free Lunch Theorem and Optimization Problems

The topic of optimization has already been extensively covered in the scientific and application literature, in forms of scientific papers as well as books. The obvious justification for our research on this topic is that most of the optimization software available works like a blackbox. However, the efficiency and robustness of the optimization algorithm (i.e. the reliability and stability of the optimal solutions obtained) are problem dependent. In essence, algorithms that are designed to be very efficient for one type of problem may not work well in another problem setting. Hence, there is no universally best algorithm for any general problem. In contrast, there is always a tradeoff between algorithmic efficiency and algorithmic robustness. This intuition is formalized by a statement called the *no free lunch theorem* in Wolpert Macready [98].

According to the *No Free Lunch Theorem*, the average efficiency of all algorithms is the same in the sense that $\sum_{L\in\Lambda} P(\hat{L}^{(n)} = \lambda|L, A_i)$ is independent of A_i, which states that one algorithm works well for one class of problems is bound to be ineffective for another class of problems.

On the other hand, the NFL theorem implies that if the problem structure is used in the algorithm design or if it is known that the algorithm and problem are well matched, then it is possible to overcome the limits of NFL in the sense that the intentionally designed algorithm can outperform substantially the blindly picking a search algorithm in available software package. Therefore our research is worthwhile.

The second reason why this research is necessary is that the past few years have witnessed an explosive development of solving optimization problem by efficient Monte Carlo simulation under the framework of Markov Chain theory. Thus the purpose of this research is, partly, to review some of these innovative ideas and bringing them together to build a powerful and efficient global optimization algorithm.

4.2 Dynamics in Complex State Space

This chapter defines the concepts concerning how to describe the state space and dynamics of a complex system. An introduction of what makes a system complex is made.

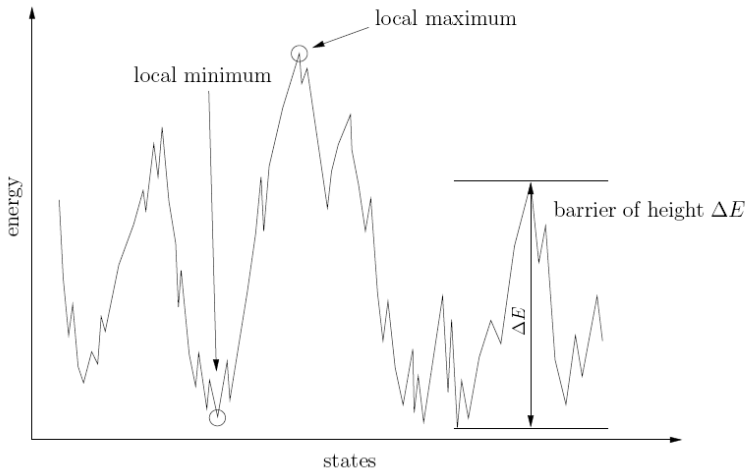

Figure 4.1: A random walk through the state space of a complex system. (Heilmann 2005) [51]

4.2.1 The State Space of a System

One of the key concepts in optimization problems is the state space of a system. It is the set of all possible states the system can be in and on this basis the dynamics can be introduced. In statistical physics, it is said that 'this dynamics introduces a probabilistic view on the evolution of a system.' [51].

Mathematically, a system is described by its state space Ω. This is simply the set of all microscopic states $\chi \in \Omega$ the system can be in. Microscopic state can be, e.g., the positions and velocities of particles in a physical system, the configuration of one model, or for example, the route plot in a *traveling salesman problem*. The number of states $L = |\Omega|$ itself can be extremely large, or even uncountable in the case of continuous space. A function $G = G(\chi)$ is defined on the states χ. It defines an 'energy for each state'. Dynamics are introduced into such systems by the concept of random walks. Random walks can be interpreted as jumps in the state space under some or no consideration. In this section we analyze several types of random walks with different behavior in a complex state space.

A local minimum is defined as a state whose energy is smaller than that of all neighboring states. Different local minima can be are separated by so-called 'high mountain ridges'. A system is called complex if the state space contains plenty of local optima (minima &

maxima). We also name these states as 'valley, hills and mountain ridges'.

4.2.2 From Random Walk to Markov Process

Because of the difficulties the deterministic optimization algorithm met, stochastic optimization procedure is introduced. In the very beginning, only a simple random walk without any steering mechanism is utilized, however, the effectiveness is not very satisfactory. Then, new types algorithms, i.e. Genetic Algorithm (GA) and Simulated Annealing (SA), were developed and widely used in solving stochastic optimization problems. However, the behavior of these stochastic optimization algorithms is quite complex and not well-understood.
So that it is helpful to understand and analyze the stochastic optimization in general on a unified theoretical ground. The Markov process formalism introduced by Heilmann 2005 [51] opens such possibility. Before going to the theory, some basic notation and definition have to be given. For a discrete state space, the move class for a specific state x is defined as $\mathcal{M}(x) \in \Omega$. The probability of selecting a new state y is denoted as $\prod_{y|x}$. Whether this random walk will move from state x to $y \in \mathcal{M}$ is determined by the so-called acceptance probability $P_{y|x}(t)$, which might be dependent on the time t. It's easy to see that the total transition probability Γ is

$$\Gamma_{y|x}(t) = \prod_{y|x} \cdot P_{y|x}(t)$$

Such transition can be seen as an element of the Markov transition matrix.

DEFINITION 4.2.1 (Markov Transition Matrix)**.** *A matrix $\Gamma \in \mathbb{R}^{n\times n}$ consists of transition probability*

$$\Gamma_{ij} = \mathbf{Prob}(X(t+1) = j | X(t) = i), \quad i, j = 1, \ldots, n.$$

The matrix must satisfy

$$\Gamma \geq 0, \qquad , P \times 1 = 1$$

where the inequality $\Gamma \geq 0$ means elementwise, i.e. $\Gamma_{ij} \geq 0$ for $i, j = 1, \ldots, n$, and 1 denotes a vector of all ones.

Because there is no rejection mechanism in the GA, so that the diagonal terms of the

transition matrix associated with the GA are all zero. While for SA algorithm, the resulting transition matrix has none zero diagonal terms, and the lower the temperature the smaller the difference between the diagonal terms and one.
Markov processes subject to different rejection rules will result in different transition probability and thus possess different dynamics in the state space. In the following section, we are going to study the relationship between the performance of the Stochastic optimization algorithm and their rejection rules.

4.3 Stochastic Optimization as Markov Process

Here in this section, we are trying to understand the SA algorithm by seeing it as a special case of Markov process and then propose a new algorithm with desirable dynamics and hence improved performance.

4.3.1 Annealing Type Algorithms

Back in 1983, Kirkpatrick [59] presented a physically motivated stochastic optimization scheme called *Simulated Annealing*. Then a family of annealing type optimization algorithms have been developed and widely applied in scientific computing.
The optimization process constructed is a mimic of a cooling process of a solid body. It consists of two steps:

1. Raise the temperature of the heat bath so that the solid melds.

2. Decrease the temperature of the heat bath slowly so that the particle can settle down in the ground state(state of lowest energy) of the solid.

At high temperature states, a solid is fluid and particles can move around freely. This ensures that a particle can move out of the attraction of local optimum. Pictorially, we can think of SA as an imaginary bounce ball jumping up and down along the landscape of the optimization landscape – from mountain to mountain, valley to valley. The higher the temperature, the more energy the ball has, so that it moves freely. At lower temperature, it is harder for the ball to move upwards to higher energy states, and tends to be attracted in a valley with low energy state. In contrast, a gradient method would only direct the ball downwards to the closest valley. For the problem with many local optima,

gradient methods tends to get stuck in a local optimum and requires several random start of the search.

The algorithm can be described by a Metropolis-Hastings dynamics. Define a state space Ω with a Boltzmann distribution $\pi(x) = \exp\left(\frac{-G(x)}{T}\right)$ (Liu 2001)[69]. The T is a pseudo temperature which scales the Boltzmann distribution. By drawing a (conditional) dependent sample x' from an easy to sample proposal distribution (usually we choose Gaussian) $q(x'x)$, we calculate the ratio $\alpha = \frac{\pi(x')q(xx')}{\pi(x)q(x'x)}$ and accept the movement from x to x' with probability $P_{x'x} = \min\{1, \alpha\}$. Intuitively, the MH will construct a random walk that will go down the energy hill for sure and go up the hill with a certain probability.This random walk will not be fixated in any local minima. The higher the temperature, the higher the probability of the algorithm can move out of the local minima.

THEOREM 4.3.1. *The random walk constructed by Metropolis-Hastings algorithm with acceptance probability to the new state x'*

$$P_{x',x} = \min\{1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\}$$

converges to a stationary distribution $\pi(x)$ (Liu 2001)[69].

The proof can be found in Appendix B.

REMARK. In practice we choose the proposal density q to be normal distribution. Because the normal distribution is symmetric, then $q(x|x') = q(x'|x)$ holds. So that the proposal density q drops out from the above equation whenever the proposal distribution is chosen to be symmetric.

> **Algorithm 2** (Simulated Annealing)**.** *Initialize a configuration (parameter setting) the temperature level T_0 and x_0, where x is a d dimensional vector.*

Step 1. *For each k (temperature level), run N_k iterations with auxiliary function $\pi_k(x) = \exp\{-\frac{G(x)}{T_k}\}$ as target function, and accept each movement by a probability of $P_{x_{n+1}|x_n} = \min(1, \frac{\pi_k(x_{n+1})}{\pi_k(x_n)})$.*

Step 2. *Pass the best configuration to the next iteration with lower temperature, $T_{k+1} \leq T_k$.*

Step 3. *Increase k to $k+1$.*

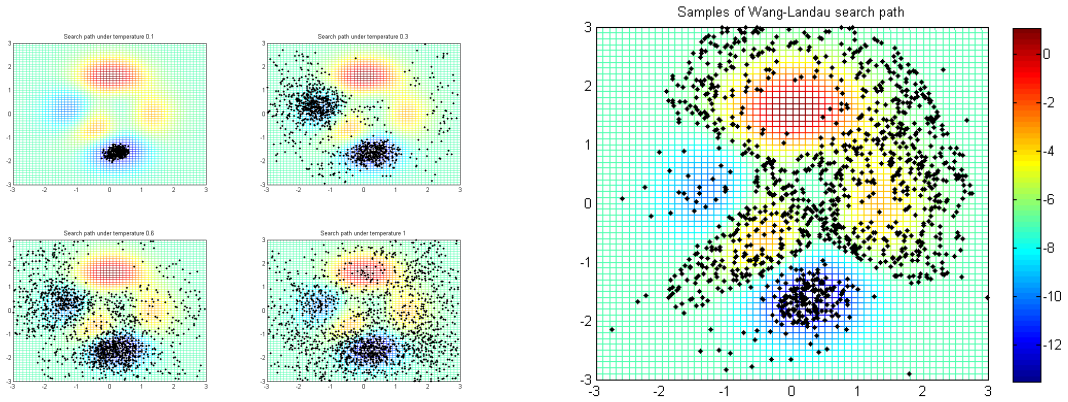

Figure 4.2: Left: The random walk constructed by Simulated Annealing (Metropolis Algorithm) under different temperatures, each of which has sample size 5000. Under high temperature, the movement is dynamic, however, also many iterations are wasted in searching uninterested region (i.e. flat plain). The right figure demonstrates the search path of the WL algorithm with 5000 iterations as well. The search is relatively uniformly distributed over the curved landscape while have few dots on the flat plain.

4.4 Beyond Simulated Annealing – Wang-Landau Algorithm

SA algorithm has been increasingly recognized by the scientific community and practitioners as an effective tool to deal with complex optimization or statistical estimation problems (Robert and Casella) [81]. Metropolis-type algorithm, the building block of the SA algorithm, allows one to break a complex task into a series of small manageable pieces. However, it is shown that its moves tend to get trapped in a local mode (Liu, Liang & Wong) [67]. A brute-force to make such algorithm less local is to increase the step size (variance of the multi-variate normal random variable) of the MH algorithm. However, this results in a very low acceptance ratio, thus an inefficient implementation.

In the past few years, the researchers were trying to develop a 'smart' global optimization algorithm which is capable of escaping out of the local 'entrapment'. There are two actively pursued direction for improvement. One approach uses temperature adapted version of simulated annealing (Kirkpatrick) [59], like simulated tempering by Geyer & Thompson [47], evolutionary Monte Carlo by Liang & Wong [63]. Another branch of efforts goes to the state space partitioning along the objective function $G(x)$ and the

weights are probably readjusted in each partition (Berg & Neuhaus) [10]. However, the major difficulty in this approach is the correct weights are a-priori unknown, so that the weights have to be estimated by a pilot simulation. The breakthrough occurs when Wang and Landau [96] proposed an efficient algorithm which can 'estimate the weights and solve the optimization problem in one breath' (Atchade & Liu) [4]. The algorithm has been successfully used to solve complex sampling problems in physics (Wang & Landau [96], Shell et al. [87], Shell et al. [88], Zhou and Bhatt [99], Liang [64]).

The idea can be described as follows: Each time the random walk hits one energy partition, the memory is kept by increasing the counter (also called the histogram) of this energy partition by one (see Figure 4.3). Then the weight associated with that energy partition $g'(e)$ is also updated by a gaining factor $g'(e)^{(k+1)} = (1+\gamma)g'(e)^{(k)}$ (the e indicates the eth sub-region). Denote the probability of the random walk hit an arbitrary state of energy E is $\pi(E)$, then the following holds

$$p(E) = \sum_{\forall\chi:G(\chi)=E} \pi_\chi = g(J(E)) \cdot \sum_{\forall\chi:G(\chi)=E} \frac{\pi_\chi}{g(J(E))}$$
$$\Rightarrow \frac{p(E)}{g(J(E))} = \sum_{\forall\chi:G(\chi)=E} \frac{\pi_\chi}{g(J(E))} = c \quad \forall E$$

where $J(x)$ is a function returns the energy ladder e of state x or a energy level such that $J : G(x_i)$or $E \to e \in \{1, 2, 3...\}$.

It's easy to see that we will have substantially higher probability of accepting movements from microscopic state x to y when $G(x) > G(y)$ than that of SA. Thus the overall acceptance rate is higher in WL algorithm.

Statement of Wang-Landau Algorithm

This statement is a slightly generalized version of Wang-Landau algorithm (Wang, Landau) [96]. Recall that in simulated annealing, we accept the transition of for a probability of $\min\left(1, \frac{\pi(x')}{\pi(\mathbf{x})}\right)$, where $\pi(\mathbf{x}) = \exp\{-\frac{G(x)}{T_k}\}$. In this case, we drop the pseudo temperature, such that $\pi(x) = \exp\{-G(x)\}$. When constructing transition probability, we use the weight associated with each sub-region $g(e)$ to rescale the $\pi(x)$, s.t.
$r^{(n)} = \min\left(1, \frac{\pi(x_{n+1})\cdot g^{(n)}(J(x_n))}{\pi(x_n)\cdot g^{(n)}(J(x_{n+1}))}\right)$.

However, this quantity $g(e)$ can not be obtained before running the algorithm, so we are

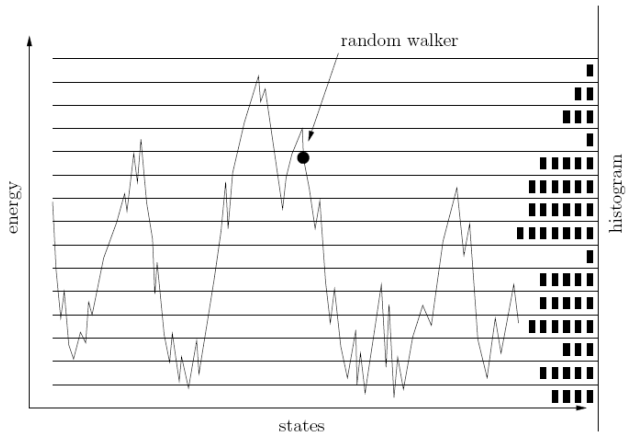

Figure 4.3: The showcase of energy partition histogram. The counter for each partition will increase 1 after been hit by the random-walk. (Heilmann) [51]

using the working estimate of $g(e)$: $g'(e)^{(k)} = g'(e)^{(k-1)} \times (1 + \gamma^{(k-1)})$ in the iteration, and updated it at each stage by a gaining factor $\gamma^{(k)} = \frac{k}{\max(k_0,k)}$, where k_0 is a pre-specified quantity we usually choose it between $2-4$ times of the number of energy partition. Of course the gaining sequence $\gamma^{(k)}$ can be chosen otherwise. However, for our optimization purpose different choice of this sequence didn't affect a lot on the final solution, because of the multi-level structure here in this case.

Wang-Landau Algorithm Setting

- Boltzmann distribution $\pi(x)$: A general non-negative function defined on sample space Ω with $\int_\Omega \pi(x)dx < \infty$, s.t. $\pi(x) = \exp\{-G(x)\}$;
- Sample space partition: $E_1, \ldots, E_m, \bigcup_{i=1}^m E_i = \Omega, E_i \cap E_j = \emptyset \quad \forall i \neq j$;
- $g(e)$ weights of each sub-region e;
- $g'(e)^{(k)}$ working estimates of the weight $g(e)$ at iteration k;
- $\gamma^{(k)}$ gaining factor at iteration k, defined as $\gamma^{(k)} = \frac{k_0}{\max(k_0,k)}$.

Mathematically, in a compact state space Ω, the idea of WL algorithm can be described as follows, let $(E_e)_{1\leq e\leq m}$ be a partition of Ω. We want to construct a random walk such that could spend the same time in each E_e and such that moves inside each E_e are done according to a standard Metropolis-Hastings algorithm with a stationary distribution π.

Note that the density for each sub-region is $g(e) := \pi(E_e)$. Then the algorithm recursively reweight π in partition E_e by a factor $g'(e)^{(n)}$ so that at the limit $g'(e)^{(n)} \propto \pi(E_e)$. Relating back the Metropolis Hastings construction that we saw before, it can be described as using X_n and some unnormalized weights $g'(e)^{(n)}$ to sample X_{n+1}, the stationary distribution is proportional to $\sum_{i=1}^{m} \frac{\pi(\mathbf{x})}{g'(i)^{(n)}} \mathbf{1}_{E_i}(x)$ ($\mathbf{1}$ is the indicator function) is used to draw X_{n+1}. Then $g'(e)^{(n)}$ is updated to $g'(e)^{(n+1)} = g'(e)^{(n)}(1 + \gamma^{(n)} \mathbf{1}_{E_e}(X_{n+1}))$. So only the weight of the sub-region that just been visited will be updated. Intuitively, $g'(e)^{(n)}$ can be seen as a multiplicative occupation measure of the set E_e, therefore as the chain stays in E_e, $g'(e)^{(n)}$ will increase exponentially increasing the probability of accepting a move outside of E_e. The comparison between conventional Markov Chain implementation (SA) and WL algorithm is shown in Figure 4.2, where we can see more global movement in search trail of WL algorithm while MH majorally consists of local moves.

Algorithm 3 (Wang-Landau algorithm)**.** *Let's start from $X_0 = x_0$ and initialize the partition-wise gaining factor $g_1^0 = g_2^0 = \cdots = g_d^0 = 1$ where the d denotes the number of partition. Define rescaled objective function (Boltzmann distribution)*
$\pi(\mathbf{x}) = \exp\big(-G(\mathbf{x})\big)$ *as the target distribution.*

Step 1. Draw sample x_{n+1} from distribution $\mathcal{N}(x_n, \Lambda \cdot I_{dim})$ where I is the identity matrix Λ is the variance coefficient.

Step 2. Calculate transition probability $\Gamma = \min\left(1, \frac{\pi(x_{n+1}) \cdot g^{(n)}(J(x_n))}{\pi(x_n) \cdot g^{(n)}(J(x_{n+1}))}\right)$; draw U from uniform distribution $[0, 1]$.

Step 3. Accept the move if $U < r$.

Step 4. Update the gain factor $g^{(n+1)}(e) = g^{(n)}(e)(1 + \gamma^{(n)} \mathbf{1}_{E_e}(X_{n+1}))$.

REMARK. The chain X_n is not a Markov chain but a controlled Markov chain controlled by the process $g'^{(n)}$. On the other hand, $g'^{(n)}$ can be seen as a stochastic approximation algorithm driven by the process X_n.

Asymptotic Behavior of WL Algorithm

The following lemma guarantees the asymptotic stability of the solution obtained by Wang-Landau.

THEOREM 4.4.1 (Convergence of Wang-Landau Algorithm (Atchade) [4])**.** *Consider the stochastic process generated by algorithm 3. Then the following holds.*

1. *for any* $i \in \{1, \ldots, d\}, g^{(n)}(i) \to g(i)$ *from arbitrary starting point almost surely as* $n \to \infty$

2. *For any* $i \in \{1, \ldots, d\}$*:*

$$\frac{1}{n}\sum_{k=}^{n-1} \mathbf{1}_{\sum_i}(X_k) \to \frac{1}{d} \quad as \quad n \to \infty$$

3. *Assume that* $f : \chi \to \mathbb{R}, |f| \leq V^{\frac{1}{2}}$*, is a measurable function. Then*

$$\frac{1}{n_i}\sum_{k=1}^{n} f(X_k)\mathbf{1}_{\sum_i}(X_k) \to \int_{\sum_i} f(x)\frac{\pi(dx)}{g(i)}, \quad as \quad n \to \infty$$

and for any $i \in \{1, \ldots, d\}$*, where* $n_i = \sum_{k=1}^{n} \boldsymbol{1}_{\sum_i}(X_k)$.

The proof is referred to the work of (Atchade) [4].

4.5 Partial Resampling

Readers might already notice that the proposal transition density in level 1 (WL algorithm) is a simple multivariate normal; if not subjecting to the MH acceptance rule, the WL generated samples would have led to a form of uniform random walk in the configuration space Ω. Liu [69] has shown that in most of the cases, choosing the proposal to be normal distribution, the resulting stochastic sampling algorithm is not very efficient, since this choice is too general and too 'noisy'. However improvements could be made by the partial resampling technique, which is a generlaized version of Gibbs conditional sampler, to be introduced in this section. The idea of Gibbs sampler is to construct a Markov Chain from a sequence of conditional distributions along the coordinates (Liu & Sabatti) [68]. A Gibbs step involves (i) decomposing variable into two parts, e.g. $x = (x_i, x_{[-i]})$ where $x_{[-i]} = \{x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$, and (ii) updating the component x_i by new sample drawn from the conditional distribution $\pi(.|x_{[-i]})$ (see Figure 4.4). A simple example of the partial resampling used in Gibbs sampler for the second coordinate can be illustrated as follows:

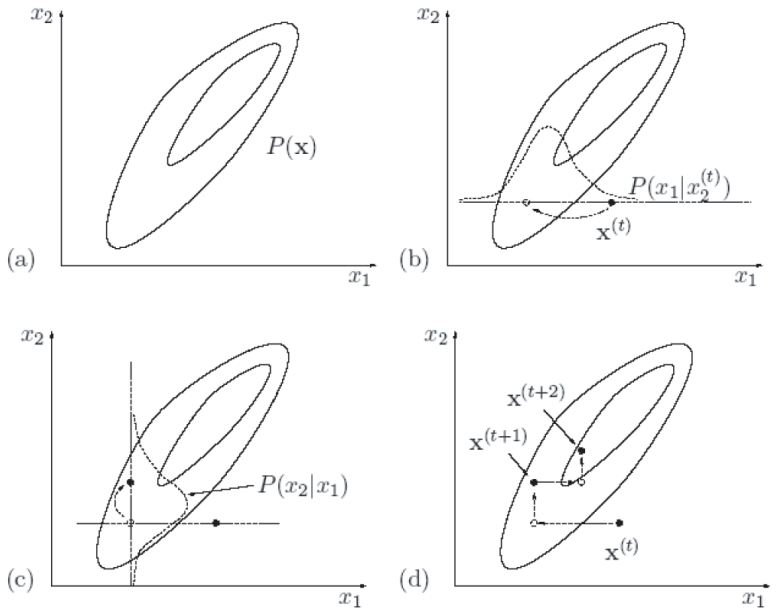

Figure 4.4: Example of Gibbs sampling in a 2D Gaussian distribution. Source: [73].

$$(x_1, x_2, \ldots, x_d) \rightarrow (x_1, x_2 + \varsigma, \ldots, x_d) \tag{4.1}$$

where ς is a random number from an appropriate distribution. Liu, Wong, & Kong [70] and Liu [71] have shown that the grouping highly correlated components together can greatly improve the efficiency of the stochastic sampler. Their conclusions are based on the findings that the Gibbs sampler's convergence rate is 'controlled by the maximal correlation between two consecutive Gibbs iterations'. It is intuitive to see that if the two consecutive samples have high correlation, the new sample adds relatively little information to the stochastic search algorithm.

The distinction between Gibbs conditional sampler and partial resampling is that the decomposition in the former one is done coordinate-wisely out of ease. However, the partial resampling is more flexible in the decomposition of the sample spaces and, more desirably, capable of incorporating problem-specific conditional distributions. In another word, the partial resampling can be seen as the generalized version of the Gibbs conditional update.

Consider the simple Gibbs update in the equation 4.1 again, the new point $x' = (x'_i, x_{[-i]})$ can be thought of as being moved from x by a transition transformation selected from the

set

$$\Upsilon = \{\varsigma \in R^1 : (x_i, x_{[-i]}) \to (x_i + \varsigma, x_{[-i]})\}.$$

The principle to choose ς is to ensure that the target distribution π is invariant with respect to the move. Therefore, a general problem arises: Suppose $x \sim \pi$ and Υ is a set of transformations. From what distribution should one draw $\varsigma \in \Upsilon$, so that $x' = \varsigma(x)$ also follows the same distribution π? Liu & Sabatti [68] gave an explicit answer to this problem when Υ forms a locally compact group. The idea is formulated in the following theorem.

THEOREM 4.5.1. *Suppose Υ is a locally compact group of transformation on Ω. Let π be a probability density, for continuous random variable, or a distribution function, for discrete random variable, for x. If $x \sim \pi(x)$ and $\varsigma \in \Upsilon$ is drawn from*

$$p_x(\varsigma)d\varsigma \propto \pi\{\varsigma(x)\}|J_\varsigma(x)|L(d\varsigma),$$

where $J_\varsigma(x) = \det\{\partial\varsigma(x)/\partial x\}$ is the Jacobian of the transformation, then $x' = \varsigma(x)$ follows π.

A proof is given by Liu & Sabatti [68]. A more flexible and problem specific construction of conditional moves in the stochastic sampling, thus, can be made by the help of this theorem. An example is to set Υ be the translation group along an arbitrary direction, i.e.,

$$\Upsilon = \{\varsigma \in R^1 : \varsigma(x) = x + \varsigma e \equiv (x_1 + \varsigma e_1, \ldots, x_d + \varsigma e_d)\},$$

where $e = \{e_1, \ldots, e_d\}$ is a fixed vector given in advance. The proper distribution for drawing ς can be derived from Theorem 4.5.1 as $p_x(\varsigma) \propto \pi(x + \varsigma e)$. Because of the the presence of the correlation among parameters, it is desirable to group the highly correlated components together. With the generalization provided by the Theorem 4.5.1, we can select ς from a random transformation group to achieve the effect of blocking/grouping: We first select a random group then update the group by drawing from the conditional distribution of a block of variables, e.g. $\pi(x_A|x_{[-A]})$. And Roberts & Sabu [80] have shown that random scan can outperform systematic scan in terms of convergence speed. Therefore, in our implementation we choose the random grouping scheme.
Practically, we first select the number of parameters we want to group at each iteration by drawing a random integer number in the range of $[1, dimension - 1]$. Then permutate this

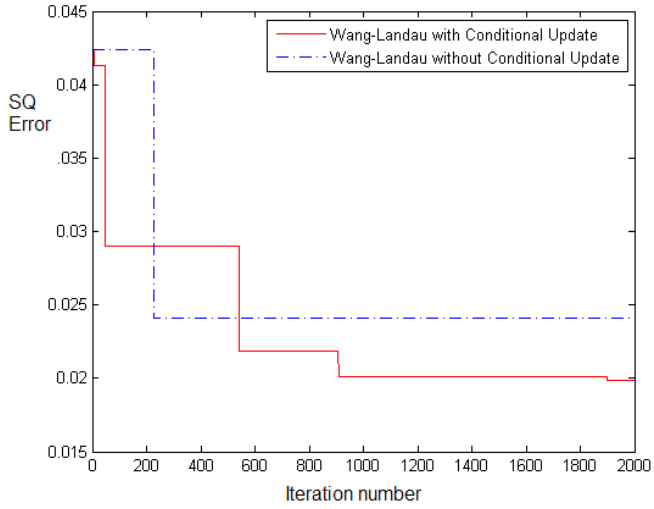

Figure 4.5: The WL with partial resampling can continuously improve the solution through, while the WL without conditional don't have any improvement after certain number of iteration, which implies a relatively low statistical efficiency.

number of particles to determine which parameters to group. Then draw a conditional multivariate normal distribution with a MH correction as the new solution. The effect of partial resampling is shown in the Figure 4.5.

During implementation, we shrink the step-size at level 2 such that it will actually refine the solution from previous level. And the energy histogram function (in the level 1) prevents the random walk from falling back to the energy region already visited. However, revisitation is not completely forbidden as the simulation time evolving when the old 'memory' is mixed with the 'new' ones. The effect of the combined effect of level 1 and level 2 is shown on a 1-D toy energy landscape of Rastrigin function (Figure 4.6).

4.6 Multi-Level HSAS Algorithm

From optimization point of view, an algorithm should be able to overcome two difficulties in complex optimization problems in general: 1) The entrapment of local minima, 2) Insufficient exploration of system minima. Our answer to overcome these difficulties is a multi-level hybrid approach: In the first level, we construct a stochastic approximation algorithm (based on Wang-Landau) to search the global energy landscape broadly with less statistical accuracy. Usually, several competing energy valleys can be detected after

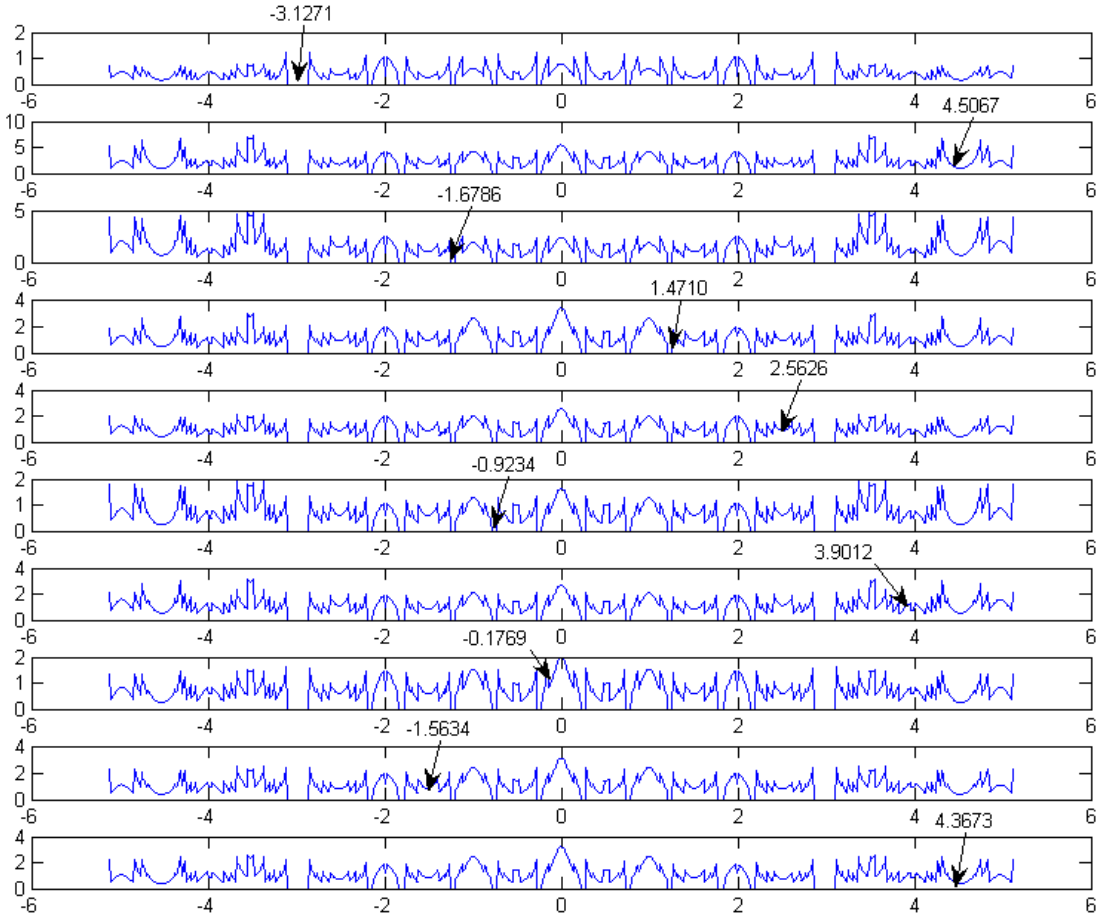

Figure 4.6: The illustration of the Wang-Landau algorithm on the landscape of Rastrigin function. The plot shows how the algorithm is transforming the energy landscape during the optimization process adaptively, while the location of system minima is preserved. The random walk is started in one local minima, the point of $x = -3$. The first row shows the transformed landscape after 1000 iteration. After being stuck in this minima for many times, the minima of $x = 3$ has been effectively flatten out. Then the algorithm is able to escape out the local entrapment and visit under visited region (second row). Each visit flats the local energy landscape and increase the probability of moving to the neighborhood region. In the last several thousands iterations, corresponding to the last two rows, the energy landscape is kept stable, and the significance of global minimum is preserved.

this procedure. In the next level, grouping the highly correlated parameters and perform a partial resampling to further explore certain sub-space where the joint effect of parameters occurs. Last but not least, a deterministic search algorithm is utilized to search inside the 'most probable' energy valley, to locate the true global optimum deep in the bottom of the valley.

Because of the nature of this algorithm is a combination of stochastic approximation and local deterministic search, so we name it to be *Hybrid Stochastic Approximation Search*, HSAS in short. The rest part of this section is devoted to the detailed construction and implementation of such algorithm.

After the Stochastic Approximation algorithm finds the probable region of global optimum, we then apply a deterministic search to track down the real optimum. The method of choice is the Nelder-Mead Simplex Method.

4.6.1 Nelder-Mead Simplex Method

Nelder-Mead Simplex Method has long been a popular non-linear optimization algorithm in practice. It based on the concept of a geometric object, simplex (in 2D, it is a triangle; in 3D, it is a pyramid, which is the *convex hull* of $d+1$ points in $\mathbb{R}^d$ not in the same supporting hyperplane. Note that this does not related to the simplex method of linear programming.

The basic idea here is that at each iteration, a new point is generated in or near the current geometric object. This new point may replace one of the current simplex vertices, yielding a new simplex. To determine the location of this new point, a reflection step is introduced where the new point is chosen to be the reflection of the worst existing point. By doing so, the simplex is moving away from the high energy landscape to the low energy landscape. The algorithm stops if the simplex is small enough. Then the solution is a point inside of the simplex. The implementation is taken from Spall [93].

Algorithm 4 (Nelder-Mead Simplex Method (Spall [93])). ***(Initialization)*** *Generate an initial set of $d+1$ vertex $\{X^d\} = (x_1, x_2, \ldots, x_{d+1})$ in $\mathbb{R}^d$ adjacent to the initial guess x_0. Calculate $G(x_i)$ for each vertex. Set values of algorithm coefficients $\alpha, \beta, \gamma, \delta$. Default settings are from Nelder and Mead's original paper, where they specify*
$\alpha = 1.0, \beta = 0.5, \gamma = 2.0, \delta = 0.5$.

Step 1. ***(Reflection)*** *Identify the vertices where the maximum, second highest, and*

minimum $G(x_i)$ happens. Let $\bar{x}$ denote the mean of all vertices x_i except $x_{\max}$. Generate a new candidate vertex x_{re} by reflecting $x_{\max}$ through $\bar{x}$ according to $x_{re} = (1+\alpha)\bar{x} - \alpha x_{\max}(\alpha > 0)$.

Step 2a. ***(Accept reflection)*** *If $G(x_{\min}) \leq G(x_{re}) < G(x_{2\,\max})$, then x_{re} replaces $x_{\max}$ in the simplex and proceed to step 3; else to step 2b.*

Step 2b. ***(Expansion)*** *If $G(x_{re}) < G(x_{\min})$, then the reflection is expanded according to $x_{exp} = \gamma x_{re} + (1-\gamma)\bar{x}$, where the expansion coefficient $\gamma > 1$; else, go to step 2c. If $G(x_{exp}) < G(x_{re})$, then x_{exp} replaces $x_{\max}$ in the simplex; otherwise, the expansion is rejected and x_{re} replaces $x_{\max}$. Go the step 3.*

Step 2c. ***(Contraction)*** *If $G(x_{re}) \geq G(x_{2\,\max})$, then the simplex contracts to reflect the poor x_{re}. Consider two cases: (i) $G(x_{2\,\max}) \leq G(x_{re}) < G(x_{\max})$ (outside contraction) or (ii) $G(x_{re}) \geq G(x_{\max})$ (inside contraction). The contraction point is determined by $\bar{x} = \beta x_{\max\,/re} + (1-\beta)\bar{x}, 0 \leq \beta \leq 1$, where $x_{re/\,\max} = x_{re}$ in case (i) or $x_{re/\,\max} = x_{\max}$ in case (ii). In case (i), if $G(x_{cont}) \leq G(x_{re})$, the contraction is accepted. In case (ii), if $G(x_{cont}) < G(x_{\max})$, the contraction is accepted. If the contraction accepted, replace $x_{\max}$ with x_{cont} and go to step 3. If the contraction is not accepted, go to step 2d.*

Step 2d. ***(Shrink)*** *The contraction has failed and the entire simplex shrinks according to a factor of $0 < \delta < 1$, retaining only $x_{\min}$. This is done by replacing each vertex x_i (except $x_{\min}$) by $\delta x_i + (1-\delta)x_{\min}$. Go to step 3.*

Step 2. ***(Termination)*** *Stop if convergence criterion is satisfied or if the maximum number of function evaluation has been reached; else, return to step 1.*

The Figure 4.7 visualize the steps of reflection, expansion, contraction, and shrinking after the failed contraction. Observe that the new vertex always lies on the line connecting $x_{\max}$ and $\bar{x}$. And the shape of simplex undergoes significant change during an expansion or contraction with standard coefficient, such that it can move itself to the low energy valley.

Advantages: Popular and effective in wide range of problems. Significant decreasing of objective function in first few iterations. Robust against noise.

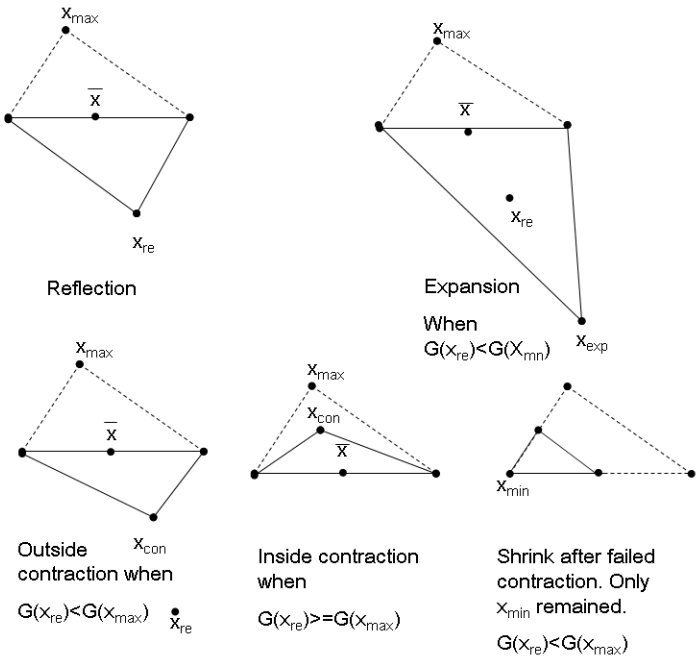

Figure 4.7: Nelder-Mead simplex method works through reflecting, expanding or contracting the $d+1$ geometrical object. The original triangular is in dashed lines.

Disadvantages: Unconstraint search. Also trapped by local optimum, require random restart.

Putting pieces together, we have the algorithm described below,

Algorithm 5 (the Hybrid Stochastic Approximation Search (HSAS) Algorithm)**.** ,

Part 0 (Definitions) *define re-scaled objective function (Boltzmann distribution)* $\pi(\chi) = \exp\big(-G(\chi)\big)$ *as the target distribution, where* G *is the object function, and* $J(x)$ *is a function returns the energy ladder* e *of state* x *such that* $J : G(x_i) \to e \in \{1, 2, 3...\}$

Part 1 (Initiation) *Choose initial guess* $X_0 = x_0$*, take* k_0*, choose* d*- the number of partitions and initialize the partition-wise gaining factor* $g_1^0 = g_2^0 = \cdots = g_m^0 = 1$*, take* $G_{best} = G(x_0)$ *and choose number of samples (iterations)* N*.*

Part 2 (Stochastic Approximation)

1. *Draw sample* x_{n+1} *from distribution* $\mathcal{N}(x_n, \Lambda \cdot I_{dim})$ *where* I *is the identity matrix* Λ *is the variance coefficient*[1]*.*

2. *Evaluate object function* $G(x_{n+1})$ *and find appropriate ladder level* $e(x_{n+1}) = J(G(x_{n+1}))$*.*

2. *Calculate transition probability* $\Gamma = \min\left(1, \frac{\pi(x_{n+1}) \cdot g^{(n)}(e(x_n))}{\pi(x_n) \cdot g^{(n)}(e(x_{n+1}))}\right)$ *and draw* U *from uniform distribution, i.e.* $U \sim \mathcal{U}([0, 1])$*.*

3. *If* $U < \Gamma$ *then accept the move and take* $G_{best} = G(x_{n+1})$*.*

4. *Update the gain factor* $g^{(n+1)}(e) = g^{(n)}(e)(1 + \gamma^{(n)})$*, where* $\gamma^{(n)} = \frac{k_0}{\max(k_0, n)}$*.*

Part 3 (partial resampling)

5. *select the number of parameters we want to group at each iteration by drawing a random integer number in the range of* $[1, dimension - 1]$*. Then permutate this number of particles to determine which parameters to group.*

6. *Draw sample* x'_{n+1} *from distribution* $\mathcal{N}(x_{n+1}, I_{new} \cdot \Lambda)$*.*

7. *Calculate sub-transition probability* $\Gamma_{sub} = \min\left(1, \frac{\Pi(x_{n+1})}{\Pi(x_n)}\right)$ *and draw* U *from uniform distribution, i.e.* $U \sim \mathcal{U}([0, 1])$*.*

8. *If* $U < \Gamma_{sub}$ *accept the movement* $x_{n+1} = x'_{n+1}$ *and update gaining factor* $g^{(n)}(e)$

8. *Check stopping criteria. If none of them satisfied, increase the index* $n + 1$ *to* $n + 2$ *and go back to step 1.*

Part 4 (Deterministic Search) *After the stopping criteria reached, perform a Nelder-Mead Simplex Method on the best solutions so far.*

[1] see below how to choose proper Λ

During implementation, there is two question of great practical importance: 1) the choice of variance coefficients Λ and 2) the choice of state space partition. These two remaining issues are discussed below.

4.6.2 The Choice of Variance Coefficient

The variance coefficient is crucial for the performance of HSAS since it determines the average step size of search movements. If you choose this too small, then the algorithm will not search broadly the state space but stuck in some local region. If it is too big, then the search will bounce in the state space arbitrarily, and the convergence to the global optimum is ruined.
The criteria of choosing variance coefficients is to adjust the acceptance rate to be 0.234 which is, in most of the cases, the asymptotic optimal acceptance rate (Roberts 2001)[82]. And we choose Λ to be the $7\% \sim 8\%$ of the scale bounded by the lower bound and the upper bound. In most of the cases, choose the coefficient to be smaller will dramatically affect the acceptance rate. While change it to be a little bigger, the effect is relatively small.

4.6.3 The Choice of State Space Partition

Our choice of space partition is based on the energy partition: We slice the space by a energy ladder of 10 levels. Although the energy ladder can be chosen quite arbitrarily, there is still some obvious principles. Firstly, the number of energy ladder depends on the complexity of the problem; Our analysis showed that in high dimension problems higher number of energy levels is preferable, since it will encourage more diverse movement across partitions. The second is that the energy difference in each partition should be small enough, less than 2 in most of the cases, so that the local MH moves in each partition will have a reasonable acceptance rate. (recall that for movement in the same subregion, the effect of gaining factor $\gamma^{(n)}$ both appear upper and below the fraction sign, so their effect is canceled out, the movement is reduced to a pure MH one). By experiment, we find that create a dense energy ladder close to the sub-region of interest (lower energy) then sparser ladder for bigger energy level work pretty fine. Because of the property that the Wang-Landau algorithm will, asymptotically, visits each subregion with equal number of times, which results in more visits to the low energy subregions.

Chapter 5

First Numerical Experiments with HSAS Algorithm

Here we evaluate the performance of the HSAS algorithm by solving several hard optimization test problems. Each test consists of 100 runs of the algorithm starting from a random initial point. We list the average, minimum, and maximum number of function evaluations needed to solve the problem. The global optimum of all test problems is known, so that the quality of the HSAS algorithm can be verified by checking the number of runs for which HSAS reaches the optimum and the distance (measured in terms of standard deviation) from the true global optimum can be determined, if it does not converge. If the algorithm stagnates in a local optimum, we conclude that in that run the algorithm fails to reach the global minimum. The results below confirm the superior performance of HSAS for these test problems. Especially the robustness against noise is very satisfactory.

5.1 Rastrigin

The first test experiment is based on the Rastrigin function, which is defined as follows,

$$f(x) = 10m + \sum_{i=1}^{m}(x_i^2 - 10\cos(2\pi x_i)), \quad -5.12 \leq x_i \leq 5.12.$$

The Rastrigin function is a special case of quadratic function. After incorporating a cosine to a quadratic function, it has many local minima. The global minimum is $f(\mathbf{x}) = \mathbf{0}$ at

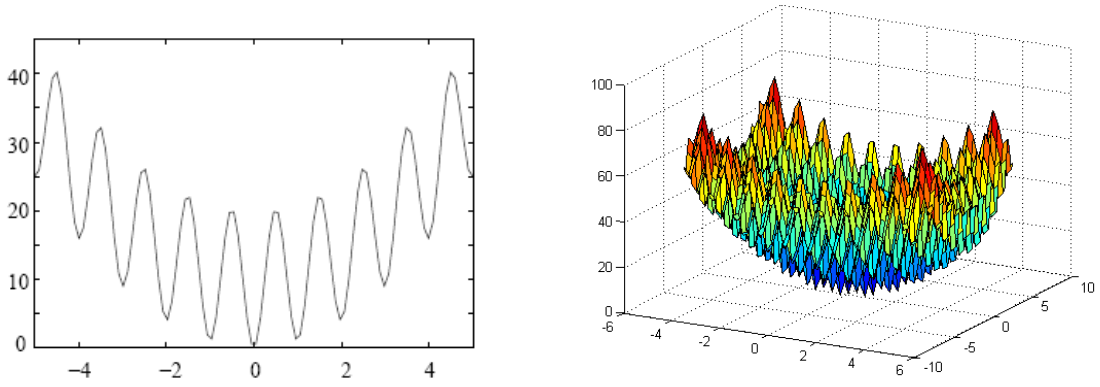

Figure 5.1: The plot of the 1D and 2D Rastrigin functions. The global optimum is located at the origin, where the function value is 0.

Table 5.1: The numerical result of HSAS over Rastrigin function

Dimension	**Number of Function Evaluation**				
	Average	Min	Max	std. dev.	fail (%)
2	3055.02	3003	3066	6.19271E-05	1%
5	4201.78	2898	12392	5.23E-3	12%
10	37729.22	19759	39163	8.33E-2	21%

$\mathbf{x} = \mathbf{0}$. The 1D and 2D versions of the Rastrigin function are shown in Figure 5.1.

The numerical experiments are done by running HSAS 100 times over the Rastrigin function in many dimensions. The HSAS algorithm is able to find the global optimum in a robust way, as almost all tests find the true optimum. Table 5.1 summarizes the results. The quality of the solution is measured by the standard deviation of the obtained numerical optima. The reason why the algorithm did not achieve absolutely zero is that the level 3 optimization (the last step with the deterministic search) stops when the stopping criterion is reached, which terminates when the radius of a triangle is less than 10^{-4}. As we can see from Table 5.1, the percentage of successful runs drops as the dimension increases. This kind of high amplitude fluctuation of the objective function is, however, uncommon in practice.

Table 5.2: The numerical convergence results of HSAS for the Schwefel function.

Dimension	**Number of Function Evaluations**			
	Average	Min	Max	fail (%)
2	1818.48	1075	3711	0
5	4201.78	2898	12392	0
10	12661.55	6972	18812	0
20	19756.81	18993	20861	0

5.2 The Schwefel Function

The difficulty of optimizing the Schwefel function comes from the fact that it has many competing minima, where each of them is separated by ridges. An algorithm capable of climbing up hills is needed and one that searches globally in the parameter space.
The Schwefel function is defined as follows,

$$f(\mathbf{x}) = \sum_{i=1}^{m}(-x_i \sin(\sqrt{|x_i|}), \quad x_i \in [-70, 70]^m.$$

The HSAS algorithm detects the global optimum in 2D version of Schwefel function $(x_1, x_2) = (65.5478, 65.5478)$ 100 out of 100 runs. The plot of some search trails, in Figure 5.2, shows that the algorithm quickly moves to the valley of interest.
For each problem, we run the optimizer 100 times. The numerical results are summarized in Table 5.2. The algorithm succeeds in finding the global optimum for all dimensions mainly because of the smoothness of the landscape. Furthermore, the standard deviation is extremely small, which means that the algorithm converges in a robust way. This result confirms the efficiency and robustness of HSAS for problems with competing global minima.

5.3 Rosenbrock Function

The performance of HSAS for the Rosenbrock function is of particular interest for this research, since it resembles closest the optimization problem in the calibration procedure, i.e., it has a unique global minimum in a long curved valley. Due to the correlation between several parameters (see Sections 3.5 and 4.5), it is expected that the optimal parameter set in the Heston model is surrounded by many similar solutions. The

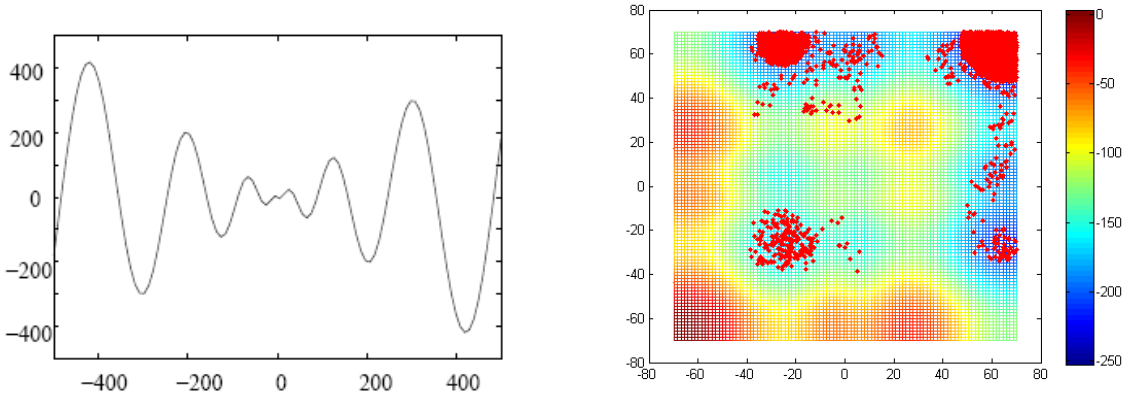

Figure 5.2: The plot of the 1-D and 2-D Schwefel functions. The 2D function domain is set to be $(x_1, x_2) \in [-70, 70] \times [-70, 70]$. The global optimum is located at a point which is close to the boundary $(65.5478, 65.5478)$. The red dots in the right figure presents the points sampled by the Wang-Landau algorithm.

Rosenbrock function reads,

$$f(x) = \sum_{i=1}^{N-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2], \quad x_i \in [-2, 2]^m.$$

It has a unique global minimum $f(\mathbf{x}) = \mathbf{0}$ for $x_i = 1, i = 1, \ldots, m$ in the valley. The 2D version is plotted in Figure 5.3. The Rosenbrock function is considered a hard problem for stochastic optimization techniques (like the genetic algorithm or simulated annealing-type algorithms). Due to the multi-level structure of the HSAS algorithm, very satisfactory convergence results are obtained for this Rosenbrock function. Here, the level 3 optimization (the Nelder simplex algorithm) contributes significantly to the convergence (level 3 can even solve this problem alone, as there is only one optimum, if the dimension of the problem is not too big, up to 5 D). Nevertheless, the HSAS profits from its multilevel structure in high-dimensional cases, as it achieves excellent convergence results and the Nelder simplex stand-alone fails in dimensions higher than 5.

The number of function evaluations and other statistics are summarized in Table 5.3. Each column represents the statistics of 100 runs. We conclude that we can obtain a fast convergence for non-trivial problems (e.g. higher dimensional ones or problems with correlation between parameters) by combining the stochastic search with the Nelder simplex algorithm.

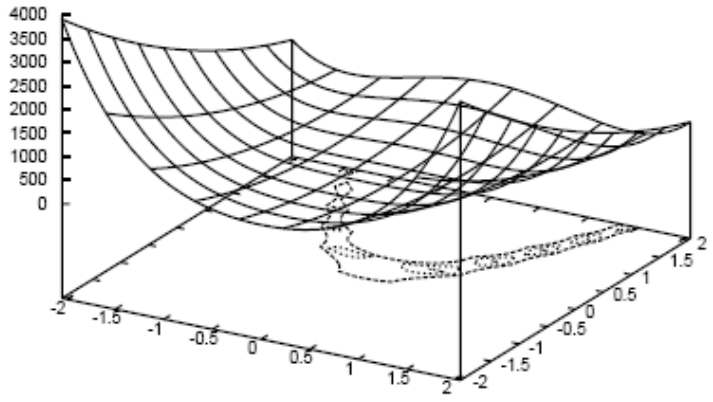

Figure 5.3: The 2D version of the Rosenbrock function. A unique global minimum $f(1,1) = 0$ is located in a long curved valley.

Table 5.3: The effect of multilevel structure on the convergence of the Rosenbrock function.

Dimension	**Number of Function Evaluation**							
	Merely Nelder Simplex				**Multilevel**			
	Average	Min	Max	fail(%)	Average	Min	Max	fail(%)
2	143.01	74	205	0%	1110.37	1060	1183	0%
5	763.24	230	1494	26%	3143.75	1910	5274	0%
10	4056.63	1381	7402	60%	19624.94	5841	29038	11%
20	16504.67	8042	25001*	100%	72921.34	64170	85000	67%

* *Note that the maximum number of function evaluations allowed is 25,000. The number of 25,000 in the last row of the table suggests that the optimizer failed at reaching the global optimum with certain runs.*

Table 5.4: The standard deviation of numerical solution of HSAS for the noisy Rosenbrock function.

Noise Amplitude	0.1	0.25	0.4	0.55	0.7	0.85	1
Nelder Simplex at level 3	0.0985	0.1914	0.1723	0.2539	0.3242	0.3121	0.3358
Quasi Newton at level 3	0.1161	0.1286	0.1498	0.1631	0.1751	0.1990	0.2176

5.4 The Rosenbrock Function with Noise

Because of the presence of model error and noise in the input (see Section 3.2), the objective function obtained by projecting real market prices of plain vanilla instruments onto the Heston model shows a non-smooth and discontinuous behavior. We assume that the non-smoothness of the objective function is of a simple shape but subject to *high-frequency, low-amplitude noise*.

We construct a two-dimensional *noisy* Rosenbrock equivalent function by adding uniform random numbers $u \sim \mathcal{U}\left([-\frac{1}{2}A, \frac{1}{2}A]\right)$ (where A denotes the amplitude) to mimic the effect of high-frequency, low-amplitude perturbations onto the Rosenbrock function. We use the HSAS solver in order to test its robustness against noise. In comparison, we also build a quasi-Newton solver at the level 3 to see the difference in performance. After testing 7 amplitudes, 0.1, 0.25, 0.4, 0.55, 0.7, 0.85, 1, respectively, the statistics of the 50 runs are summarized in Table 5.4. On this test function, the purely deterministic algorithm substantially fails at reaching global optimum, while the HSAS is still able to recover the original global optimum. Further tests also show that, when the amplitude of noise exceeds 1, the landscape of the objective function has been altered too much, that no algorithm is able to recover the true solution, as can be seen from the right-side figure of Figure 5.4.

When the noise is smaller than 0.1, then quasi-Newton method combined with the stochastic solver yields solutions that are at least as accurate as the HSAS, but converges faster. However, numerical experiments suggest that applying the combined quasi-Newton method in the Heston calibration problem will not finds better solution than HSAS. So, we assume that the noise level in the Heston calibration problem lies in the region of $[0.1, 1]$ and favor the Nelder simplex implementation.

As expected, there is a clear relationship between the noise applied on the objective function and the standard deviation (from the true global optimum $(1, 1)$) of the solution found by the optimizer . This relationship is illustrated in Figure 5.4. We can see that the standard deviation of the numerical optimum increases as the amplitude of the noise

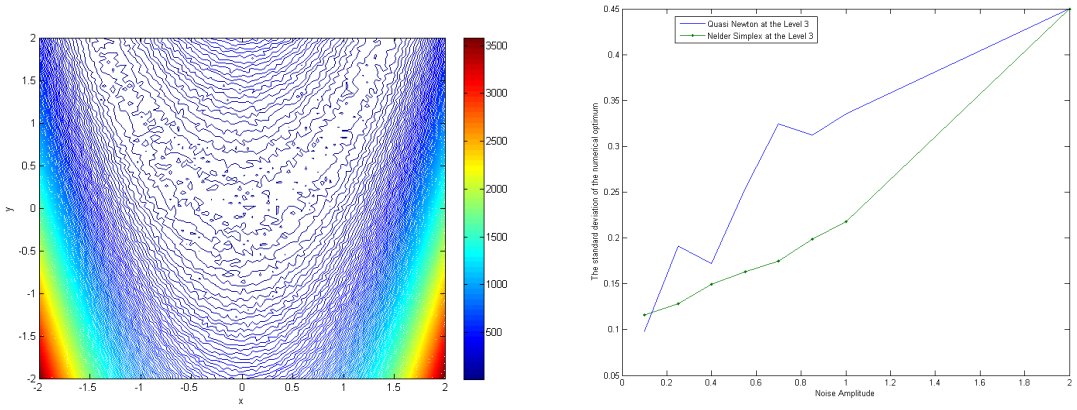

Figure 5.4: On the left: The relationship between the noise and the standard deviation of the optimum obtained from HSAS from the true global minimum of Rosenbrock. On the right: the noisy Rosenbrock function obtained by adding high frequency and low amplitude uniform random number onto the Rosenbrock function.

increases. However, it doesn't grow as fast as the noise itself. Actually, the solution remains stable under the low amplitude noise.

In contrast, the Quasi Newton method is more sensitive to the noise, because the presence of noise makes the estimation of the derivatives by finite difference schemes inaccurate. Obviously, this result holds for this specific objective function Rosenbrock and it is not possible to make any general statement just by this result. Nevertheless, we can see that the robustness of Nelder simplex algorithm (the level 3 of HSAS) against noise is quite impressive. This is one of the reasons to choose Nelder simplex as the level 3 solver of HSAS.

5.5 Recover Predetermined Parameters

A final test that has to be performed is to recover some predetermined parameters of the Heston model by first generating option prices. If the algorithm is able to find the prescribed parameters accurately, one can conclude that HSAS is effective and well-suited for solving the calibration problems occurring in the financial industry.

This is done in the following way. We choose the Heston parameters to be $\xi = (0.14, 0.421, 0.521, -0.721, 0.0234)$ and generate a 32×20 (32 strikes range from 40% to 240% and 19 maturities from 0.041yr to 14.960yr) option price matrix based on these

Table 5.5: The standard deviation (from the true parameter of Heston) of the numerical optimum after 20 runs

	κ	θ	σ	ρ	v_0
std dev.	2.34E-3	5.59E-3	3.37E-4	1.16E-3	1.74E-4

Table 5.6: Summary of 20 optimization runs

# average fcn evaluation	min # fcn	max # fcn	running time (C++)	avg. SQ Error
4734.556	4432	5146	956.417 sc	6.407E-08

parameters. We perform 20 runs of the HSAS algorithm over this price matrix to recover the Heston parameters. The standard deviation (from the true parameter set) of the numerical solution after 20 runs is chosen as the measure of robustness of the HSAS algorithm. As can be seen from Table 5.5, the HSAS is able to recover the true Heston parameters accurately.

5.6 Conclusion and Discussion

The above numerical experiments confirm the robustness of the HSAS algorithm. The robustness is due to the Nelder simplex algorithm, which is recommended for optimizing noisy functions, see Fletcher pp.18 [39] and Powell pp.230 [76]. However, it is not possible to prove convergence of Nelder simplex theoretically. Nevertheless, in this work we find it more important to have an algorithm which is effective and robust in practice.
During the research, we learned that the Sequential Quadratic Programming algorithm (implemented in Intel Math kernel library as dtrnlspbc routine) can be even more efficient for the square error minimization. It usually converges well, and requires fewer than 200 iterations to find a global optimum. However, at each step, a gradient has to be approximated by the finite difference scheme. Small variations Δx around some x then do not reflect the local behavior of the function, but the noise instead. A naive implementation of finite differences would give incorrect results. Nevertheless, gradient-based algorithms can still be applied but a sophisticated way to obtain derivatives must be used. One suitable algorithm is a technique called implicit filtering by Gilmore & Kelley [48]. This algorithm makes certain assumptions about the noise (Gaussian or uniformly distributed low amplitude), and requires a high number of function evaluations to estimate the gradient. We however favor the HSAS in this work.

Chapter 6

Applications of Stochastic Volatility Model

Due to the uncertainty of the future volatility a perfect replication of complex derivatives can not be achieved, therefore a trader needs to dynamically hedge the volatility risk exposure by trading liquid derivatives, in addition to the underlying stocks. In this chapter, we first describe the practical hedging strategy suggested by the Heston model, and then apply this model to price two volatility sensitive complex derivatives. The impact of adding stochastic volatility on the pricing is assessed in the sequel.

6.1 Practical Hedging in Heston Model

A successfully implemented hedging strategy can help investors increase the expected returns, e.g. by including private equity in a portfolio, without excessive risks. In this section, we present a real-life case of portfolio hedging and risk management under the Heston model. In order to hedge a derivative or a trading strategy involving multiple derivatives, we first determine the model parameters by calibrating the Heston model to the corresponding instruments, then evaluate the payoff and compute the hedging ratios. Because of the parametric nature of the Heston model, the hedging ratio can be easily computed: One chooses suitable reference instruments and computes the sensitivities (measured in partial derivatives w.r.t the state variables) of the overall portfolio with respect to these reference instruments. The remaining (unhedged) sensitivity is the risk exposure of this portfolio.

6.1.1 Model and Greek Notation

We again consider the Heston model. The Greeks are the quantities representing the market sensitivities of options or other derivatives. Each Greek measures a different aspect of the risk in an option position, and corresponds to a parameter on which the value of an instrument or portfolio of financial instruments is dependent. This name is used because the parameters are often denoted by Greek letters, like Delta and Gamma [57]. However, other 'Greeks', especially useful in the Heston model are

Greeks

- Vega: $\frac{\partial V_t}{\partial \sigma_t}$.
- Volga: $\frac{\partial^2 V_t}{\partial \sigma_t^2}$.
- Vanna: $\frac{\partial^2 V_t}{\partial \sigma_t \partial S_t}$.

REMARK. Readers may have already noticed that in the Heston model we specify the instantaneous variance, v_t, as the state variable, but here we talk about the volatility, σ_t. Because $v_t = \sigma_t^2$, the following holds, $\frac{\partial V_t}{\partial \sigma_t} = \frac{\partial V_t}{\partial v_t}\frac{\partial v_t}{\partial \sigma_t} = 2\frac{\partial V_t}{\partial v_t}$. Therefore, we have

- $\frac{\partial V_t}{\partial v_t} = \frac{1}{2}$Vega,
- $\frac{\partial^2 V_t}{\partial v_t^2} = \frac{1}{4}$Volga,
- $\frac{\partial^2 V_t}{\partial v_t \partial S_t} = \frac{1}{2}$Vanna.

6.1.2 P&L of a delta-hedged portfolio with Stochastic Volatility

Herein we consider the Profit and Losses, P&L, of a delta-hedged portfolio consisting of an option and a 'delta' amount of the underlying stocks. The P&L of the portfolio in the period $[t, t+dt]$ is

$$d\Pi_t = dV_t - \Delta dS_t + (\Delta S_t - V_t) r dt, \tag{6.1}$$

(with a long option and a short stock, the last term represents the remaining cash). Assume now that the stock price process follows the Heston dynamics. After applying the Taylor expansion for the V_t term, we obtain

$$dV_t = \frac{\partial V_t}{\partial t}dt + \frac{\partial V_t}{\partial S_t}dS_t + \frac{\partial V_t}{\partial v_t}dv_t + \frac{1}{2}\left[\frac{\partial^2 V_t}{\partial S_t^2}dS_t^2 + \frac{\partial^2 V_t}{\partial v_t^2}dv_t^2\right] + \frac{\partial^2 V_t}{\partial S_t \partial v_t}dS_t dv_t.$$

Inserting the above identity into equation 6.1, we get,

$$d\Pi_t = \frac{\partial V_t}{\partial t}dt + \frac{\partial V_t}{\partial S_t}dS_t + \frac{\partial V_t}{\partial v_t}dv_t + \frac{1}{2}\left[\frac{\partial^2 V_t}{\partial S_t^2}dS_t^2 + \frac{\partial^2 V_t}{\partial v_t^2}dv_t^2\right] + \frac{\partial^2 V_t}{\partial S_t \partial v_t}dS_t dv_t$$
$$- \Delta dS_t + r\Delta S_t dt - \underbrace{\left(\frac{\partial V_t}{\partial t} + \frac{1}{2}\sigma_{\text{impl}}^2 S_t^2 \frac{\partial^2 V_t}{\partial S_t^2} + rS_t\frac{\partial V_t}{\partial S_t}\right)}_{=rV_t dt \text{ because of BS PDE}} dt.$$

In the underbraced term, the V_t is the market price of the option at time t. It is obtained by substituting the implied volatility into the Black-Scholes PDE. Rewriting the above equation, we find

$$d\Pi_t = \frac{1}{2}S_t^2\left[\sigma^2 - \sigma_{\text{impl}}^2\right]dt + \frac{\partial V_t}{\partial v_t}dv_t + \frac{1}{2}\frac{\partial^2 V_t}{\partial v_t^2}dv_t^2 + \frac{\partial^2 V_t}{\partial S_t \partial v_t}dS_t dv_t \tag{6.2}$$
$$= \underbrace{\frac{1}{2}S_t^2\left[\sigma^2 - \sigma_{\text{impl}}^2\right]}_{\text{Gamma P\&L}}dt + \underbrace{\frac{\partial V_t}{\partial v_t}dv_t + \frac{1}{2}\frac{\partial^2 V_t}{\partial v_t^2}\eta^2 v_t dt + \frac{\partial^2 V_t}{\partial S_t \partial v_t}\rho\eta S_t v_t dt}_{\text{P\&L sensitivity to volatility exposure}} \tag{6.3}$$

where ρ, η are the parameters specified in the Heston model.
Using the Greeks notation to write out the partial derivatives, we get the following:

$$P\&L^{Vol} = \frac{1}{2}\text{Vega } dv_t + \frac{1}{8}\text{Volga } \eta^2 v_t dt + \frac{1}{2}\text{Vanna } \rho\eta S_t v_t dt \tag{6.4}$$

In the Heston model the exposure of the portfolio value to volatility changes is analytically tractable, accurately up to second order. As we can see from Figure 6.1, the volga and vanna have a non-trivial impact on the portfolio value. Once the higher order volatility sensitivity is controlled in the Heston model, the Heston model is well-suited to robustly price and hedge volatility sensitive derivatives.

6.2 Pricing Complex Derivatives

Because of the existence of a semi-analytical Fourier transform solution (based on the characteristic function) for the Heston model, the prices of a wide class of payoffs can be computed efficiently and accurately. The characteristic function can express option prices in terms of Fourier integrals, which are numerically evaluated by the FFT (described in Appendix A). The option pricing formula for a general payoff is discussed in Lee [61]. In

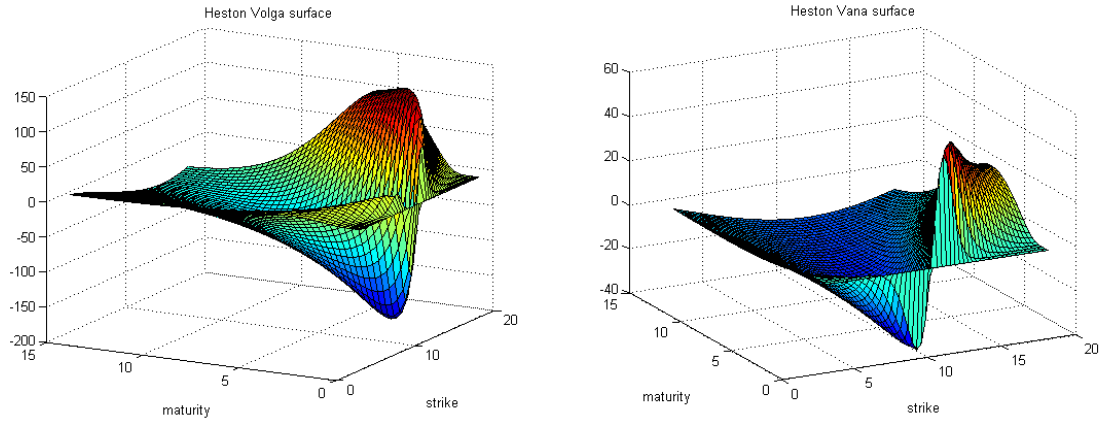

Figure 6.1: The left side figure shows the Heston volga surface which represents the second order sensitivity (derivatives) w.r.t. variance v_t. The right side figure depicts the Heston vanna surface which is the cross derivative of stock, s, and variance, v_t.

this section, we only focus on pricing issues of some volatility sensitive exotic options. The complex derivatives considered are:

EXAMPLE 6.2.1 (Forward Start Calls)**.** This contract has the following payoff:

$$\text{Payoff} = \left(\frac{S_{t_2}}{S_{t_1}} - K\right)^+,$$

for $0 < t_1 < t_2 < T$ and $K \geq 0$.

Parameter K denotes the strike price.

EXAMPLE 6.2.2 (Globally Floored Cliquets)**.** This contract is determined by the following payoff:

$$\text{Payoff} = \left\{ \sum_{i=1}^{d} \min \left(\text{LocalCap}, \max \left(\frac{S_{t_i}}{S_{t_{i-1}}} - 1, \text{LocalFloor} \right) \right) \right\}.$$

where the d denotes the resetting period of this contract. This contract will be explained in some more detail in the following sections.

EXAMPLE 6.2.3 (Variance Swap)**.** This contract has the following payoff:

$$\text{Payoff} = \frac{252}{N} \sum_{k=1}^{N} \left(\log \frac{S_{t_k}}{S_{t_{k-1}}} \right)^2 - K$$

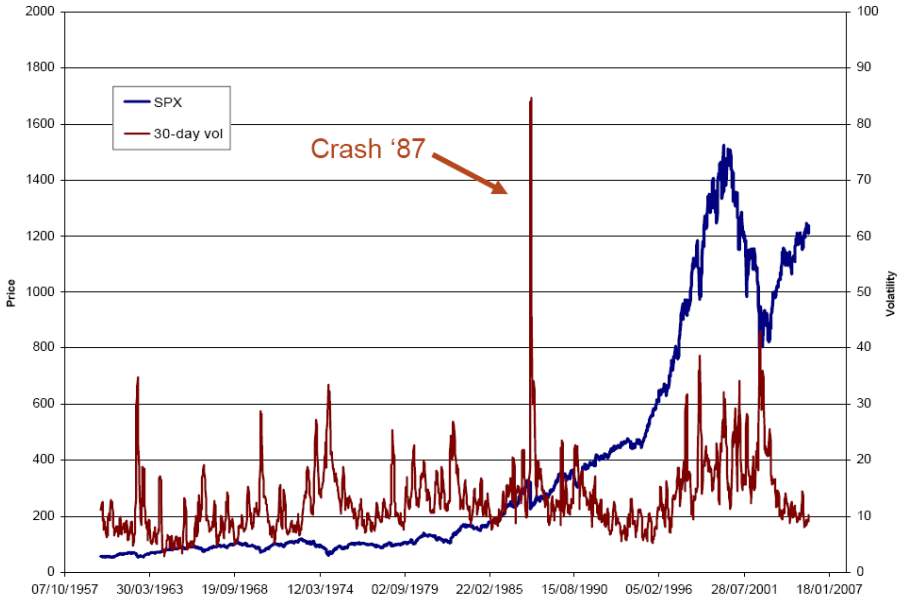

Figure 6.2: S&P 500 Spot level and 30-day realized volatility. There is a negative correlation between the spot movement and the volatility. It is natural to use a variance swap as a hedge for market downward movements. Source of picture (Buehler 2005) [18]

where 252 denotes the number of trading days per year, N is the number of business days until settlement.

EXAMPLE 6.2.4 (Option on Variance)**.**

$$\text{Payoff} = \left(\sum_{k=1}^{N} \left(\log \frac{S_{t_k}}{S_{t_{k-1}}} \right)^2 - K \right)^+$$

6.2.1 Variance Swap Contract

The variance swap has become more and more liquid in recent years. Its popularity comes from the fact that it enables investors to trade the volatility as an asset. Since the volatility tends to move opposite to the market, volatility rises as the market falls (see Figure 6.2), it is a natural hedge against the downward market movements.

A *variance swap* with maturity T is a contract which pays out the *realized annualized variance* of the log return up to time T, defined as $RV_T = \frac{252}{N} \sum_{k=1}^{N} \left(\log \frac{S_{t_k}}{S_{t_{k-1}}} \right)^2$, in

exchange for a fixed strike:

$$\mathbb{E}_0^Q RV_T - K = \mathbb{E}_0^Q \frac{252}{N} \sum_{k=1}^{N} \left(\log \frac{S_{t_k}}{S_{t_{k-1}}} \right)^2 - K = 0.$$

It has a zero initial value. Thus the rational price of such a contract is the expectation of the realized variance, RV_T.

The pricing of a variance swap is mathematically intriguing, since its payoff can be replicated by a portfolio of European call and put options and its rational price can be approximated by the quadratic variation of the underlying process. However, this method based on quadratic variation (described in Appendix C) relies on some strong assumptions:

A **No jumps**. Asset price S_t evolves as a continuous diffusion.

B **Perfect static hedge**. A continuum of calls and puts across all strikes can be used as hedging vehicles.

C **Continuous sampling**. A sampling procedure which is sufficiently frequent to permit the realized variance to approximate its continuous limit – which is its quadratic variation:

$$RV_T \approx \quad < \log(S_T) > \quad = \quad \frac{1}{T} \int_0^T v_t dt.$$

Here we present a solution method called the moment method, which can approximate the expected payoff of the variance swap payoff under the Heston model efficiently and allows us to drop the assumptions above.

Moment Method for Variance Swap

LEMMA 6.2.5 (Moments Representation of Variance Swap Payoff)**.** *The expectation of the realized variance can be written as follows:*

$$\begin{aligned}\mathbb{E}^Q[RV_T] &= \mathbb{E}^Q\left[\frac{1}{T}\sum_{i=1}^{N}\left[\log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right)\right]^2\right] \\ &= \frac{1}{T}\sum_{i=1}^{N}\mathbb{E}^Q[(x_{t_i}-x_{t_{i-1}})^2] \\ &= \frac{1}{T}\sum_{i=1}^{N}\mathbb{E}^Q[x^2_{t_i,t_{i-1}}] \\ &= -\frac{1}{T}\sum_{i=1}^{N}\frac{\partial^2\phi_{t_{i-1},t_i}(u)}{\partial u^2}\bigg|_{u=0}\end{aligned}$$

where $x_t = \log(S_t)$ and $x_{t_i,t_{i-1}}$ denotes the log return in the time interval t_i, t_{i-1}. It is the second moment of the log stock return.

Recall that the characteristic function can be used as a moment generating function – taking the second derivative of the characteristic function gives the second moment. However, the technical complexity is that the starting point of each time interval is forward in time, so that we have to use the *forward characteristic function*, $\phi_{t^*,T}(u)$, we derive in Appendix C.2.1 instead of the usual one. Taking the second order derivative around the point $u = 0$, we obtain the second moment of the stock log-return under the Heston model:

$$\frac{\partial^2\phi_{t^*,T}(u)}{\partial u^2}\bigg|_{u=0} = i^2\mathbb{E}^Q[x^2_{t^*,T}] = -\mathbb{E}^Q[x^2_{t^*,T}].$$

After adjusting the time interval $\{t^*, T\}$ to be the variance swap discrete sampling interval $\{t_{i-1}, t_i\}, 1 \le i \le N$, we have,

$$\begin{aligned}\mathbb{E}^Q[RV_T] &= \frac{1}{T}\sum_{i=1}^{N}\mathbb{E}^Q[x^2_{t_i,t_{i-1}}] \\ &= -\frac{1}{T}\sum_{i=1}^{N}\frac{\partial^2\phi_{t_{i-1},t_i}(u)}{\partial u^2}\bigg|_{u=0}.\end{aligned}$$

Table 6.1: The impact of increasing the jump intensity (in rows) and its size (in columns) on the expectation of the realized variance. The realized variance estimated by quadratic variation approximation, without jumps, is 0.0139. APJS means 'average percentage of the jump size'

	jump intensity (λ)			
APJS (μ_J):	0.05	0.1	0.3	0.6
-0.05	0.0139* (100%)	0.0144 (103.6%)	0.015 (107.9%)	0.0155 (111.5%)
-0.1	0.0144 (103.6%)	0.15 (107.9%)	0.0172 (123.7%)	0.0205 (147.5%)
-0.3	0.0205 (147.5%)	0.0266 (191.4%)	0.052 (374.1%)	0.0902 (648.9%)
-0.6	0.0559 (402.1%)	0.0979 (704.3%)	0.266 (1913.6%)	0.518 (3726.6%)

Since we also have the characteristic function for Bates model, evaluating the second moment of the Bates dynamics is easily computed. It gives the risk-neutral expectation of the realized variance when incorporating both diffusion and jump. The impact of including jumps is presented in Table 6.1. Technical details are discussed in Appendix C. We see in the table that the jump size (in percentage) has a first order effect on the expectation of the realized variance, whereas the jump intensity (measuring the frequency of jumps) has a relatively smaller impact on the expected realized variance. Because the market is regulated in such a way that all trades are suspended if jumps over 5% occur in the market, we only have to consider the first row in Table 6.1 carefully. Under this extreme market situation, if jumps arrive at a high intensity, we need to consider an extra (around 10%) jump risk premium when establishing a hedge with variance swaps (see first row, last column of the table). Under the usual market conditions, it is safe to use the quadratic variation approximation to price and calculate the hedge ratios.

6.2.2 Cliquet Type Structure

Here, we consider the 3 years Global Floored Cliquet which has the Dow Jones EURO STOXX50 index as the underlying and offers a guaranteed principal redemption plus an annual coupon to be paid on December 2nd of each year, given by

$$\max\left\{\sum_{i=1}^{36}\min\left(1\%,\max\left(\frac{S_{t_i}}{S_{t_{i-1}}}-1,-1\%\right)\right),\text{MinCoupon}\right\}$$

where the MinCoupon is 2% and the annual coupon is capped at 12% (12 times -1 %). To price cliquets, we start with basic insights. Note that we can decompose the payoff into a

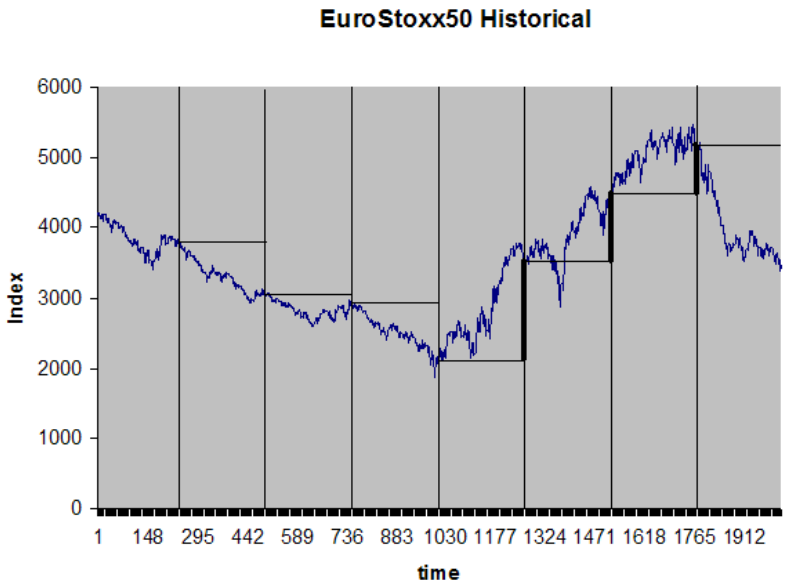

Figure 6.3: Illustration of a cliquet payoff. The payoff of the cliquets is the relative size of stock price movement between two resetting dates. The thick solid lines represents nonzero cliquet payoffs.

sum of 'legs' paying $\frac{S_{t_i}}{S_{t_{i-1}}} - 1$, which are called the ATM forward starting options (see Figure 6.3). Intuitively, the behavior of such a payoff should behave like the spread of a portfolio of ATM forward starting calls, despite the global floor. We also expect that this structure is very sensitive to assumptions made on the forward skew. The local volatility model (the industrial standard smile-consistent model which assumes a deterministic volatility) tends to underprice this structure since the forward skew given by this model is too flat (see Figure 6.4). While the forward skew in the Heston model is getting steeper as the starting time moves further into the future. The price of forward starting options in the Heston model can also be valued by the Fourier transformation technique, as presented in Appendix A. The technical implementation is described in the sections C.2 and C.2.1. We can see in Figure 6.4, that the computation confirms our intuition that the Heston model price is higher than the one with the local volatility model.

We also find that this product is sensitivity to the vol-vol parameter. The reason is that when the volatility is high, the actual payoff hits the LocalCap and thus will not change anymore. At this time, this structure has a zero vega, but when the volatility decreases, the vega sensitivity increases. Therefore, a writer has to sell volatility when it is cheap

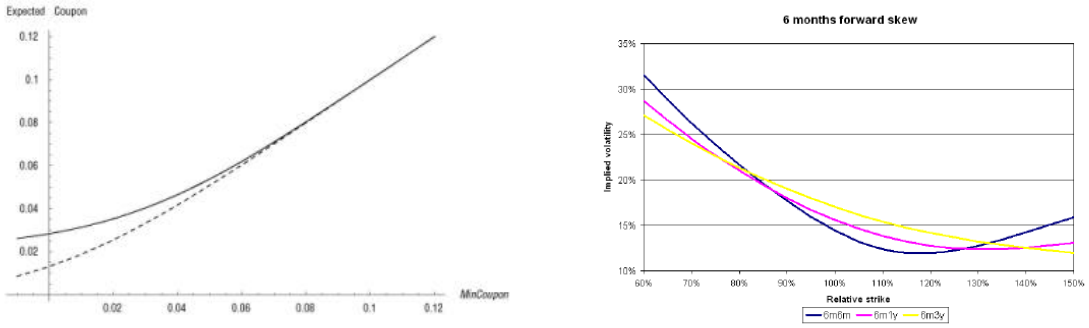

Figure 6.4: Left side figure shows the model price of a locally capped and globally floored cliquet as a function of MinCoupon. The solid line represents the price under stochastic volatility, the dashed line under local volatility. The local volatility model underprices this product [44]. The right side figure shows that the 6 months implied forward skew is steeper when the starting time of the contract moves forward (we see a 6 month contract that starts either after 6 months (blue), or after 1 year (red) or after 3 years (yellow).

(the level is low) and buy volatility when it is expensive to neutralize the trading book in terms of vega. One has to make a good estimation of how many times the volatility profile shifts significantly over the maturity of the hedge. The cost of being wrong is significant. In a stochastic volatility model, the vol-of-vol parameter is a measure of such a shift of the volatility profile.

6.3 Conclusion

In this chapter, we presented some applications of the Heston model when pricing complex derivatives. Because the state space in the Heston model includes the instantaneous variance, v_t, the Heston model is capable of characterizing the sensitivity with respect to the volatility levels. Therefore, this model suits the task of analyzing and managing trading strategies with a substantial exposure to volatility risks well.
We have demonstrated the use of the Heston model for pricing complex derivatives. The major advantage of using this model lies in the existence of a semi-closed-form solution based on the Fourier transform for European style payoffs. The prices of the complex products can be obtained in a fraction of a second, as long as the payoff can be replicated by payoffs of European style products. The two cases demonstrated here, i.e. the pricing of a cliquet and of a variance swap, the product prices can be computed within 1 second.

Chapter 7

Conclusion

As mathematicians and students tend to be overwhelmed by the magic of mathematics, we have to remind ourselves that mathematical models, especially those dealing with stochastic processes, of course, are 'incorrect' in some ways'. The interesting question is not which one is fully correct but whether a model is useful in the sense that it can give practitioners practical recommendations for his relevant questions.

Because of the complexity of the modern financial derivatives, like nowadays option contracts, practitioners heavily rely on mathematical models to price the derivatives. The first chapter described the problem setting and the validity of underlying assumptions. We discussed the risk associated with the financial derivatives and formulated the calibration problem. It is, however, well-known that the classical Black-Scholes option pricing model is not capable of pricing without a significant bias. Numerous model extensions have been introduced in recent years and in this thesis we have discussed several of them. To improve the understanding of the dynamics of the future uncertainty of asset prices, in particular, the stochastic volatility models, especially the Heston model, have emerged as a new industrial standard in the domain of exotic equity derivative. Therefore the work in this thesis was devoted to this model.

In Chapter 2 we presented the analysis of the implied volatility in Black-Scholes models. Stochastic volatility and jumps in asset prices are the essential features for mathematical models to construct models with features that are consistent with market data. This advanced model however has a serious drawback: it contains several parameters that need to be fitted to the present financial market data. Any model for pricing financial derivatives is useful only if we can combine it with the 'correct' parameters in the sense

that market data can be reproduced well and stable over time.
In the industry, one way of calibration is the so-called 'implied' methodology for derivative pricing. Calibration boils down to solving an inverse problem: First one assumes a parametric model which governs the underlying asset price process, then one tries to find the parameters such that the difference between the prices generated by the model and the financial market prices is smallest. Chapter three then presented a numerically stable and efficient calibration procedure for the calibration of all parameters and for fast daily recalibration. The approach introduced was based on a least squares formulation which is solved by a multi-level hybrid stochastic optimization algorithm with also a deterministic search component. In this thesis we have developed a calibration procedure to find optimal parameters. The algorithm consisted of several steps:

1. Evaluation of the quality of data by a novel 'check on arbitrage-free data'. By this check we could filter out the erroneous data.

2. Implementation of an efficient and robust global optimization algorithm (HSAS). We successfully produced a high-quality fit for the Heston model and its jump extension to the market prices.

3. Allow for some time-dependent parameters to improve the calibration maturity-wise.

Chapter four gave a detailed description of the multilevel-structure global optimization algorithm, named the Hybrid Stochastic Approximation Search (HSAS) algorithm, which is the main contribution in this thesis. It is a multilevel-structured global optimization algorithm which contains a stochastic approximation, for robustness, and a deterministic search, for efficiency. In the stochastic approximation we employed two Markov processes with different dynamics in hierarchy. The first level is called the Wang-Landau algorithm, whose direction was guided by information from previous iterations by a sample space partitioning function. This scheme is effective in helping the algorithm escaping from the local entrapment, but the side effect is that the search steps may vary too much in the state space. Therefore, the algorithm is not very efficient in finding the global optimum. The second level, the partial resampling, served as the solution refining step. By constraining the algorithm in certain subspaces, at each iteration, the partial resampling step enhanced the convergence. A local deterministic search is finally performed. As a result, the algorithm performs efficiently and effectively on the optimization test problems (with multiple local minima) and the Heston model calibration problem.

Numerical results for optimizing some test functions and a model calibration based on true Heston parameters were presented in Chapter 5.

The HSAS algorithm was also capable of handling the jump extension of the Bates model. The jump component in the asset price model adds extra quality to a model in the sense that it generates more skewed volatility surfaces at short maturity end. The calibration procedure could find a set of stable time independent parameters to fit the full European option market price matrix (including the extreme short end skew). Applying this calibration procedure to a time-series of market data, we found that the correlation and vol-vol parameters fluctuated in time, while the other parameters were more stable. The algorithm has shown to be particularly robust, especially for problems with, say, up to ten parameters to optimize.

It should be mentioned that there is still plenty of room for improving the performance of the optimization algorithm, because its asymptotic convergence behavior is not fully understood and requires further investigation.

Chapter six then presented financial applications of the Heston model for pricing and managing some exotic derivative securities, like the variance swap or the cliquet option. Portfolio hedging was also outlined in some detail.

The extension of the calibration method to multi-asset portfolios is left as another very interesting research direction. A multi-disciplinary approach guided by mathematical reasoning, inspired by experience of financial engineers, and supported by the information in a wealth of financial data is probably the best way to perform this research.

Appendix A

Integral Transform Solution of Option Price

In order to perform the calibration procedure, in this section, we gives details of a efficient algorithm based on the *Fast Fourier Transform* to evaluate European options in the Heston model. A fairly general mathematical framework is presented for obtaining option prices driven by general underlying processes and payoffs. The idea of the integral transform method is built upon the fundamental insight that the option price can be obtained by taking the integral of the payoff function over an appropriate pricing kernel (risk neutral density). Let us assume there exists a martingale measure, then the price of a contingent claim $C_0(x,t)$ with payoff $\phi(x_T)$ is given by the expectation:

$$\begin{aligned} C_0(x,t) &= \mathbb{E}_t^Q[e^{-\int_t^T r(s)ds}\phi(x_T)] \\ &= e^{-\int_t^T r(s)ds}\int q(x_T,T;x_t,t)\phi(x_T)dx_T \end{aligned}$$

The pricing kernel is found by solving a (non-homogeneous) ODE obtained by the integral transform (Laplace or Fourier) of the forward Kolmogorov equation.

A.1 Portfolio Replication in the Presence of Stochastic Volatility

Here we extend the portfolio replication argument to the case of stochastic volatility. Under the Heston model, we set up a portfolio Π, containing the option being priced whose value is denoted by $C(S, v, t)$, a quantity $-\Delta$ of the stock and a quantity of $-\Delta_1$ of another asset which has the same source of randomness as the volatility process, C_1. We have

$$\Pi = C - \Delta S - \Delta_1 C_1$$

The change in this portfolio in a time interval, dt, is given by

$$\begin{aligned} d\Pi = &\{\frac{\partial C}{\partial t} + \tfrac{1}{2}vS^2(t)\frac{\partial^2 C}{\partial S^2} + \rho\eta v\beta S(t)\frac{\partial^2 C}{\partial v \partial S} + \tfrac{1}{2}\eta^2 v\beta^2\frac{\partial^2 C}{\partial v^2}\}dt \\ &- \Delta_1\{\frac{\partial C_1}{\partial t} + \tfrac{1}{2}vS^2(t)\frac{\partial^2 C}{\partial S^2} + \rho\eta v\beta S(t)\frac{\partial^2 C_1}{\partial v \partial S} + \tfrac{1}{2}\eta^2 v\beta^2\frac{\partial^2 C_1}{\partial v^2}\}dt \\ &+ \underbrace{\{\frac{\partial C}{\partial S} - \Delta_1\frac{\partial C_1}{\partial S} - \Delta\}}_{\text{Random part 1}} dS \\ &+ \underbrace{\{\frac{\partial C}{\partial v} - \Delta_1\frac{\partial C_1}{\partial v}\}}_{\text{Random part 2}} dv \end{aligned}$$

By arbitrage-free constraint, we formulate a delta hedging strategy such that the randomness from the asset and the volatility processes cancels. More precisely, we set the following coefficients to be zero:

$$\frac{\partial C}{\partial S} - \Delta_1\frac{\partial C_1}{\partial S} - \Delta = 0, \tag{A.1}$$

$$\frac{\partial C}{\partial v} - \Delta_1\frac{\partial C_1}{\partial v} = 0. \tag{A.2}$$

This leaves us with

$$\begin{aligned} d\Pi =&\{\frac{\partial C}{\partial t}+\tfrac{1}{2}vS^2(t)\frac{\partial^2 C}{\partial S^2}+\rho\eta v\beta S(t)\frac{\partial^2 C}{\partial v\partial S}+\tfrac{1}{2}\eta^2 v\beta^2\frac{\partial^2 C}{\partial v^2}\}dt \\ &-\Delta_1\{\frac{\partial C_1}{\partial t}+\tfrac{1}{2}vS^2(t)\frac{\partial^2 C}{\partial S^2}+\rho\eta v\beta S(t)\frac{\partial^2 C_1}{\partial v\partial S}+\tfrac{1}{2}\eta^2 v\beta^2\frac{\partial^2 C_1}{\partial v^2}\}dt \\ =&r\Pi dt \\ =&r(C-\Delta S-\Delta_1 C_1)dt \end{aligned}$$

where we have used the fact that the return on a risk-free portfolio must equal the risk-free rate r which we will assume to be deterministic here. Substitute condition A.1 into the above expression to replace the Δ and Δ_1 terms Collecting all the C terms on the left side and all C_1 terms on the right-hand side, we get the following partial differential equation:

$$\begin{aligned} &\frac{\frac{\partial C}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 C}{\partial S^2}+\rho\eta vS\frac{\partial^2 C}{\partial v\partial S}+\frac{1}{2}\eta^2 v\frac{\partial^2 C}{\partial v^2}+rS\frac{\partial C}{\partial S}-rC}{\frac{\partial C}{\partial v}} \\ =&\frac{\frac{\partial C_1}{\partial t}+\frac{1}{2}vS^2\frac{\partial^2 C_1}{\partial S^2}+\rho\eta vS\frac{\partial^2 C_1}{\partial v\partial S}+\frac{1}{2}\eta^2 v\frac{\partial^2 C_1}{\partial v^2}+rS\frac{\partial C_1}{\partial S}-rC_1}{\frac{\partial C_1}{\partial v}} \end{aligned}$$

Since the left is a function of C only, and the RHS is a function of C_1, this equation implies that both sides are equal to a function independent of C and C_1. For the sake of simplicity, we choose the functional form provided by $\kappa(v-\theta)+\Lambda(S,v,t)\sqrt{v}$ which keeps the analytical tractability. The term $\Lambda(S,v,t)$ is called the market price of volatility risk. Different choices of this term result in different risk neutral measures. This renders the market incompleteness.

Now let us assume that we are in a risk-adjusted world, and the agents are indifferent to the market price of risk. Or in other words, we set $\Lambda(S,v,t)=0$ and proceed to solve the remaining PDE

$$\frac{\partial C}{\partial t}+\tfrac{1}{2}vS^2\frac{\partial^2 C}{\partial S^2}+\rho\eta vS\frac{\partial^2 C}{\partial v\partial S}+\tfrac{1}{2}\eta^2 v\frac{\partial^2 C}{\partial v^2}+rS\frac{\partial C}{\partial S}-rC=\kappa(v-\theta)\frac{\partial C}{\partial v} \tag{A.3}$$

where the option value C is the solution of this partial differential equation (Gatheral)[43].

Before solving equation A.3 with appropriate boundary conditions, we can simplify it by making some suitable change of variable. Let us write the original stochastic process in terms of the forward price $F_t=\log(e^{\int_t^T r_s ds}S(t))$. Because of the assumption of constant

interest rates we can simply taking $F_t = \log(e^{\int_t^T r_s ds} S(t)) = \log(e^{r(T-t)} S(t))$. After applying Itô's lemma we observe that the F_t is driftless under the risk-neutral measure

$$dF_t = -r_s F_t dt + r_s F_t dt + \eta F_t dW^Q = \eta F_t dW^Q$$

where W^Q denotes the Brownian motion under the risk-neutral measure.
To ease notation, we write x for F_t for the later derivation:

$$\frac{dC}{dt} = \frac{\partial C}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial C}{\partial t} = -r\frac{\partial C}{\partial x} + \frac{\partial C}{\partial t}.$$

After canceling out the $-r\frac{\partial C}{\partial x}$ term, we arrived at:

$$\frac{\partial C}{\partial t} + \tfrac{1}{2}v(\frac{\partial^2 C}{\partial x^2} - \frac{\partial C}{\partial x}) + \rho\eta v\frac{\partial^2 C}{\partial v \partial x} + \frac{1}{2}\eta^2 v\frac{\partial^2 C}{\partial v^2} - rC = \kappa(v-\theta)\frac{\partial C}{\partial v}. \tag{A.4}$$

A.1.1 Characteristic Function for the Heston Model

The PDE A.4 can be solved by taking a Fourier transform on both sides of the equation, then solve the reduced ODE with proper boundary condition. Following the derivation in the Section 2 of Gatheral [44], we start from the Fourier transformed forward Kolmogorov equation:

$$-\frac{\partial \phi(u)}{\partial \tau} - \tfrac{1}{2}u^2 v\phi(u) \pm \tfrac{1}{2}iuv\phi(u) + \tfrac{1}{2}\eta^2 v\frac{\partial^2\phi(u)}{\partial v^2} + \rho\eta iuv\frac{\partial\phi(u)}{\partial v} + (a-bv)\frac{\partial\phi(u)}{\partial v} = 0 \tag{A.5}$$

Define

$$\alpha = -\frac{u^2}{2} - \frac{iu}{2} + iu$$
$$\beta = \lambda - \rho\eta - \rho\eta iu$$
$$\gamma = \frac{\eta^2}{2}$$

then the Equation A.5 reads

$$v[\alpha\phi(u) - \beta\frac{\partial P}{\partial v} + \gamma\frac{\partial^2\phi(u)}{\partial v^2}] + a\frac{\partial\phi(u)}{\partial v} - \frac{\partial\phi(u)}{\partial\tau} = 0 \tag{A.6}$$

The detailed derivation of $\phi(u)$ is omitted here and we take the characteristic function as given.

Because of the importance of this result, we formulate this as a lemma:

LEMMA A.1.1. *For the Heston model, the characteristic function $\phi_T(u)$ conditioned on the initial value x_0, v_0 of the underlying diffusion process $x_T = \log S_T$ is defined by:*

$$\phi_T(u) = \exp(\underbrace{iu(x_0 + rT)}_{C}) \tag{A.7}$$

$$\cdot \exp\left(\underbrace{\frac{v_0}{\eta^2}\left[\frac{1 - e^{-DT}}{1 - Ge^{-DT}}\right](\kappa - \rho\eta iu - D)}_{B(u,T)v_0}\right) \tag{A.8}$$

$$\cdot \exp\left(\underbrace{\frac{\kappa\theta}{\eta^2}\left[T(\kappa - \rho\eta iu - D) - 2\log\left(\frac{1 - Ge^{-DT}}{1 - G}\right)\right]}_{A(u,T)}\right) \tag{A.9}$$

where

$$D = \sqrt{(\kappa - \rho\eta iu)^2 + (u^2 + iu)\eta^2}, \quad G = \frac{\kappa - \rho\eta iu - D}{\kappa - \rho\eta iu + D}$$

and x_0, v_0 are the initial values for the log-price and the volatility process, respectively.

REMARK. The characteristic function is essentially the Fourier transform of the density generated by underlying process (can be of diffusion or of Lévy type) conditioned on the initial state of the process. The physical variable x_t has been integrated out, leave the frequency variable u in equation.

A.2 Carr-Madan Inversion

Carr and Madan [23] presented an efficient and stable way of inverting the characteristic function to get the option price. The payoff function can be rewritten as

$$C_T(k) := (\exp(x) - \exp(k))^+$$

However, $C_T(k)$ tends to S_0 as k tends to $-\infty$, and hence the payoff function is not square integrable . To obtain a square integrable function, we have to add a damping factor to modify the call price function, where $c_T(k)$ is defined by:

$$c_T(k) \equiv \exp(\alpha k) C_T(k)$$

for a proper $\alpha > 0$. Now the Fourier transform of $c_T(k)$ is defined by:

$$\psi_T(v) = \int_{-\infty}^{\infty} e^{iuk} c_T(k) dk$$

We first obtain an analytical expression for $\psi_T(u)$ which is equivalent to taking the Fourier transform over the pricing kernel (risk neutral density), integrate out the log strike, then obtain call prices numerically using the inverse transform:

$$C_T(k) = \frac{\exp(-\alpha k)}{2\pi} \int_{-\infty}^{\infty} e^{-iuk} \psi_T(u) du = \frac{\exp(-\alpha k)}{\pi} \int_0^{\infty} e^{-iuk} \psi_T(u) du \tag{A.10}$$

Carr and Madan then found the Fourier transform of the exponential dampened option price $\psi_T(u)$ in an elegant closed form in terms of the characteristic function of the underlying diffusion process $\phi_T(u)$:

$$\begin{aligned} \psi_T(u) &= \int_{-\infty}^{\infty} e^{iuk} \int_k^{\infty} e^{\alpha k} e^{-rT} (e^x - e^k) q_T(x) dx dk \\ &= \int_{-\infty}^{\infty} e^{-rT} q_T(x) \int_{-\infty}^{x} (e^{x+\alpha k} - e^{(1+\alpha)k}) e^{iuk} dk dx \\ &= \int_{-\infty}^{\infty} e^{-rT} q_T(x) [\frac{e^{(\alpha+1+iu)x}}{\alpha + iu} - \frac{e^{(\alpha+1+iu)x}}{\alpha + 1 + iu}] dx \\ &= \frac{e^{-rT} \phi_T(u - (\alpha+1)i)}{\alpha^2 + \alpha - u^2 + i(2\alpha+1)u} \end{aligned} \tag{A.11}$$

The Carr-Madan inversion scheme is more accurate and stable than the Gil-Palaez scheme, where two complex integrals have to be calculated. For detailed explanation, the interested reader is referred to ManWo Ng [74]. Our implementation and later investigation are based on the Carr-Madan inversion scheme.

A.3 FFT Implementation of Carr-Madan Invesrion Scheme

We have obtained the characteristic function $\phi_T(u)$ in closed form, and substitute it into equation A.12 to get $\psi_T(u)$ and solve the following integral:

$$C_T(k) = \frac{\exp(-\alpha k)}{\pi} \cdot c_T(k) = \frac{\exp(-\alpha k)}{\pi} \int_0^{\infty} e^{-iuk} \psi_T(u) du.$$

Carr & Madan [23] showed how to utilize the FFT algorithm to solve this integral. Their numerical scheme is outlined in the following. Let us temporally write $g(u) = e^{-iuk}\psi_T(u)$ for ease of the notation, and perform the Trapezoidal rule on the u axis

$$c_T(k) = \int_0^\infty \underbrace{e^{-iuk}\psi_T(u)}_{g(u)} du \approx \int_0^M g(u)du \approx \frac{\Delta u}{2}\Big[g(u_1) + 2\sum_{n=2}^{N-1} g(u_n) + g(u_N)\Big] \quad \text{(A.12)}$$

$$= \Delta u\Big[\sum_{n=1}^{N} g(u_n) - \frac{1}{2}[g(u_1) + g(u_N)]\Big] \quad \text{(A.13)}$$

where $M = N\Delta u$ and $u_n = (n-1)\Delta u$. Then, Carr and Madan discretized the log-strike axis $k_m = -b + \Delta k(m-1), \text{for} \quad m = 1, \ldots, N$, to create a full Fourier matrix. Constant $b \in \mathbb{R}$ is assumed to be the lower boundary point of the log-strike axis. Further sorting out the summation terms, we have:

$$\begin{aligned} c_T(k_j) &= \int_0^\infty \underbrace{e^{-iuk}\psi_T(u)}_{g(u)} du \approx \Delta u\Big[\sum_{n=1}^{N} e^{-i[(n-1)\Delta u][-b+\Delta k(j-1)]}\psi_T(u_n) - \frac{1}{2}[g(u_1) + g(u_N)]\Big] \\ &= \Delta u\Big[\sum_{n=1}^{N} \underbrace{e^{-i\Delta k\Delta u(n-1)(j-1)}}_{\omega_N^{(n-1)(j-1)}} e^{i(n-1)b\Delta u}\psi_T(u_n) - \frac{1}{2}[g(u_1) + g(u_N)]\Big] \\ &= \Delta u\Big[\sum_{n=1}^{N} \omega_N^{(n-1)(j-1)}[e^{i(n-1)b\Delta u}\psi_T(u_n)] - \frac{1}{2}[g(u_1) + g(u_N)]\Big]. \end{aligned}$$

$$\begin{pmatrix} c_T(k_1) \\ c_T(k_2) \\ \vdots \\ c_T(k_N) \end{pmatrix} \approx \Delta u \begin{pmatrix} e^{-iu_1k_1} & e^{-iu_2k_1} & \ldots & e^{-iu_Nk_1} \\ e^{-iu_1k_2} & e^{-iu_2k_2} & \ldots & e^{-iu_Nk_2} \\ \vdots & \vdots & \vdots & \vdots \\ e^{-iu_1k_N} & e^{-iu_2k_N} & \ldots & e^{-iu_Nk_N} \end{pmatrix} \times \begin{pmatrix} \psi_T(u_1) \\ \psi_T(u_2) \\ \vdots \\ \psi_T(u_N) \end{pmatrix} - \frac{1}{2}\begin{pmatrix} g(u_1) \\ 0 \\ \vdots \\ g(u_N) \end{pmatrix}$$

$$= \Delta u \underbrace{\begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & e^{-i1\Delta u1\Delta k} & \ldots & e^{-i(N-1)\Delta u1\Delta k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & e^{-i1\Delta u(N-1)\Delta k} & \ldots & e^{-i(N-1)\Delta u(N-1)\Delta k} \end{pmatrix}}_{\text{This Fourier matrix is solved efficiently by the FFT}} \cdot \begin{pmatrix} e^{iu_1b}\psi_T(u_1) \\ e^{iu_2b}\psi_T(u_2) \\ \vdots \\ e^{iu_Nb}\psi_T(u_N) \end{pmatrix} - \frac{1}{2}\begin{pmatrix} g(u_1) \\ 0 \\ \vdots \\ g(u_N) \end{pmatrix}$$

Appendix B

Proofs

Proof of the lemma 3.5.1 term structure of Heston's BS implied volatility. Taken the Heston model of (2.2) and (2.3) with a substitution of $x_t := \log(\frac{S_t}{K})$. Consider the case under forward measure such that $\mu = 0$,

$$dx_t = -\frac{v_t}{2}dt + \sqrt{v_t}dW_t^1 \tag{B.1}$$

$$dv_t = \kappa(\theta - v_t)dt + \eta\rho\sqrt{v_t}dW_t^1 + \sqrt{1-\rho^2}\eta\sqrt{v_t}dW_t^2 \tag{B.2}$$

eliminating the $\sqrt{v_t}dW_t^1$, we have

$$dv_t = \kappa(\theta - v_t)dt + \eta\rho(dx_t + \frac{v_t}{2}dt) + \sqrt{1-\rho^2}\eta\sqrt{v_t}dW_t^2$$

From this equation, we can see that the Heston model, as well as stochastic volatility model in general, is essentially a joint process of stock price and instantaneous variance. Take expectation on both sides of this equation and consider the unconditional expectation of the instantaneous variance, $\hat{v}_s$, at time s. Solving this equation, we have

$$\begin{aligned}
\mathbb{E}\int_0^s dv_t &= \mathbb{E}\int_0^s \kappa(\theta - v_t)dt + \underbrace{\mathbb{E}\int_0^s \eta\rho\sqrt{v_t}dW_t^1}_{=0} + \underbrace{\mathbb{E}\int_0^s \sqrt{1-\rho^2}\eta\sqrt{v_t}dW_t^2}_{=0} \\
&\Rightarrow -\int_0^s \frac{1}{\mathbb{E}v_t - \theta}d(\mathbb{E}v_t - \theta) = \int_0^s \kappa dt \\
&\Rightarrow -\log(\mathbb{E}v_t - \theta)|_0^s = \kappa s \\
&\Rightarrow \hat{v}_s = \mathbb{E}v_s = (v_0 - \theta)e^{-\kappa s} + \theta
\end{aligned}$$

Then define the expected total variance at time t by the following identity

$$\hat{w}_t := \int_0^t \hat{v}_s ds = (v_0 - \theta)\left\{\frac{1 - e^{-\kappa t}}{\kappa}\right\} + \theta t \tag{B.3}$$

After taking the expectation, we note $u_t := \mathbb{E}[v_t | x_T]$ as the expectation of the instantaneous variance at time t conditional on the final value x_T.

$$du_t = -\kappa(u_t - \theta)dt + \frac{\rho\eta}{2}u_t dt + \rho\eta\frac{x_T}{\hat{w}_T}d\hat{w}_t + \sqrt{1-\rho^2}\eta\sqrt{v_t}\mathbb{E}[dW_t^2 | x_T]$$

If the dependence of dW_t^2 on x_T is weak, say $\sqrt{1-\rho^2}$ is small, we can drop the last term and approximate the previous equation by

$$\begin{aligned} du_t &\approx \left[(\frac{\rho\eta}{2} - \kappa)u_t + \kappa\theta\right]dt + \rho\eta\frac{x_T}{\hat{w}_T}d\hat{w}_t \\ &= -\kappa'(u_t - \theta')dt + \rho\eta\frac{x_T}{\hat{w}_T}\hat{v}_t dt \end{aligned}$$

where $\kappa' = \kappa - \frac{\rho\eta}{2}, \theta' = \frac{\theta\kappa}{\kappa'}$. The solution to the above equation is

$$u_T \approx (v - \theta')e^{-\kappa' T} + \theta' + \rho\eta\frac{x_T}{\hat{w}_T}\int_0^T \hat{v}_s e^{-\kappa'(T-s)ds} \tag{B.4}$$

Following Gatheral [44]'s derivation in the third chapter, the Black-Scholes Implied variance is the following

$$\sigma_{BS}^2(K,T) = \frac{1}{T}\int_0^T v_{K,T}(t)dt = \frac{1}{T}\int_0^T \mathbb{E}^{G_t}[\sigma_t^2]dt \approx \frac{1}{T}\int_0^T v_L(\hat{x}_t)dt$$

where $\{\hat{x}_t\}$ is the most probable path. Since we defined the conditional expectation $u_t := \mathbb{E}[v_t | x_T]$ which agrees with the definition of the local volatility as shown in the page 14 of [44]:

$$\sigma^2(K,T) = v_L(x_t) = \mathbb{E}[v_T | x_T = K] = u_T \tag{B.5}$$

Together with the expression of u_T we put in equation C.1 and the definition of the local volatility, we complete the equation and arrived at the following

$$\sigma^2(K,T) \approx \frac{1}{T}\int_0^T v_L(\tilde{x}_t)dt = \frac{1}{T}\int_0^T u_t(\tilde{x}_t)dt \tag{B.6}$$

$$\approx \frac{1}{T}\int_0^T \hat{v}'_t dt + \rho\eta\frac{x_T}{\hat{w}_T}\frac{1}{T}\int_0^T dt\int_0^t \hat{v}_s e^{-\kappa'(t-s)}ds \tag{B.7}$$

Setting the $x_T = 0$ in equation above and perform the integral, we have the at-the-money term structure of BS implied variance in the Heston model:

$$\begin{aligned}\sigma^2(K,T) &\approx \frac{1}{T}\int_0^T \hat{v}'_t dt = \frac{1}{T}\int_0^T \left[(v-\theta')e^{-\kappa' t} + \theta'\right]dt \\ &= (v-\theta')\frac{1-e^{-\kappa' T}}{\kappa' T} + \theta'\end{aligned}$$

which is the expression we saw in lemma 3.5.1 hence finish the proof. □

Proof of theorem 4.4.1 MH algorithm's convergence to its unique equilibrium distribution π.

LEMMA B.0.1 (Detailed balance condition)**.**

$$T(y|x)\pi(\mathbf{x}) = T(x|y)\pi(\mathbf{y}) \quad \forall x, y \in \Omega \tag{B.8}$$

for all $x, y \in \Omega$, where $T(y|x)$ are the transition probabilities of an ergodic Markov chain X. Then π defines the equilibrium distribution of X.

Proof. Given the detailed balance condition, it is easy to show that

$$\int \pi(x')T(x|x')dx' = \int \pi(x)T(x'|x)dx' = \pi(x)\int T(x'|x)dx' = \pi(x)$$

Thus we can see that in the state space Ω, the probability of jumping to a new state x' from any x equals to $\pi(x)$. Thus we call the whole chain is in equilibrium and it converges to a unique stationary distribution which is equal to the target distribution π. □

Suppose a state space Ω is continuous, and the π is the distribution we are interested in. In MH algorithm, we denote the $q(x'|x)$ for the proposal transition probability (e.g. Gaussian distribution) from x to x', and because of the reject-accept mechanism, the we

accept the proposed transition with probability

$$T(x'|x) = q(x'|x) \times \min\left\{1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\right\}$$

$\pi(x)$. We want to show that under such reject-accept scheme, the MH algorithm defines a transition probability $T(x'|x)$ satisfies the detailed balance condition equation B.8, such that the whole chain converges to π as its stationary distribution from any starting point.

Consider two states x and x'. Either

$$\pi(x)q(x'|x) < \pi(x')q(x|x')$$

or

$$\pi(x)q(x'|x) > \pi(x')q(x|x')$$

Firstly, let's assume $\pi(x)q(x'|x) > \pi(x')q(x|x')$, then we can see that $\frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)} < 1$, so the transition probability of moving from x to x' under the MH algorithm is

$$\begin{aligned} T(x'|x) &= q(x'|x) \min\left\{\underbrace{\frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}}_{<1}, 1\right\} \\ &= q(x'|x) \times \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)} \\ &= \frac{\pi(x')}{\pi(x)} q(x|x') \end{aligned}$$

$$\Rightarrow T(x'|x)\pi(x) = \pi(x')q(x|x') \tag{B.9}$$

Now $\frac{\pi(x)q(x'|x)}{\pi(x')q(x|x')} > 1$, so the backward probability of moving from x' to x is

$$\begin{aligned} T(x|x') &= q(x|x') \min\left\{\underbrace{\frac{\pi(x)q(x'|x)}{\pi(x')q(x|x')}}_{>1}, 1\right\} \\ &= q(x|x') \end{aligned}$$

Substitute the last identity into equation B.9, we have

$$T(x'|x)\pi(x) = \pi(x')T(x|x')$$

which suggests the detailed balance condition holds.
The case when $\pi(x)q(x'|x) < \pi(x')q(x'|x)$ can be proved exactly in the same light that you only have to switch the value of r (forward $r(x'|x) = 1$ and backward $r(x|x') = \alpha$). $\square$

Appendix C

Variance Swaps and Forward Starting Options

C.1 Pricing Variance Swap in the Heston Model

In the Heston model,

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^1 \\ dv_t &= \kappa(\theta - v_t)dt + \eta\rho\sqrt{v_t}dW_t^1 + \sqrt{1-\rho^2}\eta\sqrt{v_t}dW_t^2 \end{aligned}$$

we take the log transformation $x_t = \log S_t$ to the stock process, such that

$$dx_t = d\log S_t = (\mu - \frac{1}{2}v_t)dt + \sqrt{v_t}dW_t^1$$

It is known that in the limiting case when $N \to \infty$, the realized variance converges to the quadratic variation of x_t, and by the lemma C.1.2

$$\lim_{N\to\infty} \sum_{i=1}^{N} \left\{ \log S_i - \log S_{i-1} \right\}^2 = \quad < x_T > \quad = \int_0^T v_t dt$$

DEFINITION C.1.1 (Quadratic Variation)**.** *For any partition* $\pi = \{t_0 \le t_1 \le t_2 \le \cdots \le t_n\}$ *of* $[0,t] \subset [0,T]$ *and for any process* $\{X_t\}$ *on* $[0,T]$*, the*

π-quadratic variation of the process $\{X_t\}$ is defined to be the random variable

$$Q_\pi(X_t) = \sum_{i=1}^{n} (X_{t_i} - X_{t_{i-1}})^2.$$

If there is a process $\{V_t\}$ such that $Q_\pi(X_t)$ converges in probability to V_t for any sequence of partitions $\{\pi\}$ of $[0,t]$ such that $\mu(\pi) \to 0$ as $n \to \infty$, then we say that $\{V_t\}$is the quadratic variation of $\{X_t\}$. When the quadratic variation of $\{X_t\}$ exists, we denote it by $< X_t >$.

LEMMA C.1.2 (Quadratic Variation for Standard Process)**.** *For any standard process that*

$$dX_t = a(\omega, s)ds + b(\omega, s)dW_s \quad \textit{where} \quad 0 \le t \le T,$$

the quadratic variation of X_t exists, and it is given by

$$< X_t > \quad = \int_0^t b^2(\omega, s)ds \qquad \forall 0 \le t \le T$$

The proof is skipped. Interested reader is referred to (pp. 129, Steele 2001)[94]. Then applying the expectation operator over the payoff (realized variance) with the risk neutral probability measure, we obtained the expected value of variance swap contract

$$\mathbb{E}[< \log(S_t) >] = \mathbb{E}[\int_0^T v_t dt] = \int_0^T \mathbb{E}(v_t)dt \tag{C.1}$$

The last identity is because of the Fubini Theorem.

The $\mathbb{E}(v_t)$ is computed as follows:

$$\begin{aligned}
\mathbb{E}\int_0^s dv_t &= \mathbb{E}\int_0^s \kappa(\theta - v_t)dt + \underbrace{\mathbb{E}\int_0^s \eta\rho\sqrt{v_t}dW_t^1}_{=0} + \underbrace{\mathbb{E}\int_0^s \sqrt{1-\rho^2}\eta\sqrt{v_t}dW_t^2}_{=0} \\
&\Rightarrow -\int_0^s \frac{1}{\mathbb{E}v_t - \theta} d(\mathbb{E}v_t - \theta) = \int_0^s \kappa dt \\
&\Rightarrow -\log(\mathbb{E}v_t - \theta)|_0^s = \kappa s \\
&\Rightarrow \mathbb{E}v_t = (v_0 - \theta)e^{-\kappa t} + \theta
\end{aligned}$$

Substitute the last identity into Equation C.1, the expected total variance is

$$\begin{aligned}\mathbb{E}[< \log(S_t) >] &= \int_0^T \mathbb{E}(v_t)dt = \int_0^T (v_0 - \theta)e^{-\kappa t} + \theta dt \\ &= \frac{1 - e^{-\kappa T}}{\kappa}(v_0 - \theta) + \theta T\end{aligned}$$

The annualized variance can be simply calculated by dividing the total variance by T.

Therefore, we can see that the price of a variance swap contract in the Heston model only depends on the expected level of variance but does not depend on correlation ρ and volatility of variance η, since they are the coefficients of martingale differences (Brownian motion) which is vanished after taking expectations.

C.2 Forward Starting Option

Let t^* and T be a pair of maturities such that $t^* < T$. A *forward starting option* is a vanilla option with expiry T and strike αS_{t^*} set at time t^*, which is equal to αS_{t^*}. The quantity S_T is the underlying asset, but it can also be an index. More precisely, the payoff structure of this option at time T is given by

$$V_{FS}(T) = \left(\frac{S(T)}{S(t^*)} - K\right)^+ \tag{C.2}$$

Thus under risk-neutral measure, the initial time (t_0) price is

$$\begin{aligned}V_{FS}(t_0) &:= \mathbb{E}^{\mathbb{Q}}\left[e^{-rT}(\frac{S_T}{S_{t^*}} - K)^+ | \mathcal{F}_0\right] \\ &= e^{-r(t^*-t_0)} \underbrace{\mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t^*)}(\frac{S_T}{S_{t^*}} - K)^+ | \mathcal{F}_0\right]}_{\text{option price start from } t^*} \\ &= e^{-r(t^*-t_0)}\mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t^*)}(\frac{S_0 \exp[r(T-t_0) - \frac{1}{2}\int_{t_0}^T v_t dt + \int_{t_0}^T \sqrt{v_t}dW_t^1]}{S_0 \exp[r(t^*-t_0) - \frac{1}{2}\int_{t_0}^{t^*} v_t dt + \int_{t_0}^{t^*} \sqrt{v_t}dW_t^1]} - K)^+ | \mathcal{F}_0\right] \\ &= e^{-r(t^*-t_0)}\mathbb{E}^{\mathbb{Q}}\left[e^{-r(T-t^*)}(\frac{\exp[r(T-t_0) - \frac{1}{2}\int_{t_0}^T v_t dt + \int_{t_0}^T \sqrt{v_t}dW_t^1]}{\exp[r(t^*-t_0) - \frac{1}{2}\int_{t_0}^{t^*} v_t dt + \int_{t_0}^{t^*} \sqrt{v_t}dW_t^1]} - K)^+ | \mathcal{F}_0\right]\end{aligned}$$

In the Black-Scholes world where the constant volatility is assumed, this is easy. We just take the underbraced part and substitute it with an call option price of that term.

$$V_{FS}^{BS}(t_0) = e^{-r(t^*-t_0)}C(K, 1, T-t^*, \sigma_{BS})$$

where $C(K, S, \tau, \sigma)$ is the BS price of an European option of the term $t^* - t_0$. Although we all know that the constant volatility is not true, we still can use the BS forward start formula to get the *BS implied forward volatility*,

$$\sigma_{BS\_For}(t^*, T-t^*, K),$$

which is a function of strike, maturity, and forward starting time. Moreover, there are more important facts implied by the derivation above.

1. The delta of forward starting options is zero, since the S_0 term has been canceled out.

2. Forward starting options are gamma neutral (i.e. $\frac{\partial^2}{\partial S^2}V_{FS}(t_0) = 0$)

3. The contract has non zero vega. So it is a purely vega bet.

□

C.2.1 Pricing Forward Starts in the Heston Model

LEMMA C.2.1 (Forward Characteristic Function of the Heston)**.** *The forward (conditional) characteristic function for the Heston model is defined as*

$$\begin{aligned}\phi_{t^*,T}(u) &:= \mathbb{E}^{\mathbb{Q}}[\exp(iu(x_T - x_{t^*}))|x_0, v_0] \\ &= \exp\left[iur\tau + A(u,\tau) + \left(\frac{e^{-\kappa t^*}B(u,\tau)}{1-2\beta_{t^*}B(u,\tau)}\right)v_0\right]\left(\frac{1}{1-2\beta_{t^*}B(u,\tau)}\right)^{\frac{2\kappa\theta}{\eta^2}}\end{aligned}$$

where

$$A(u,\tau) = \frac{\kappa\theta}{\eta^2}\left[T(\kappa - \rho\eta i u - D) - 2\log\left(\frac{1-Ge^{-DT}}{1-G}\right)\right]$$
$$B(u,\tau) = \frac{1}{\eta^2}\left[\frac{1-e^{-DT}}{1-Ge^{-DT}}\right](\kappa - \rho\eta i u - D)$$
$$D = \sqrt{(\kappa - \rho\eta i u)^2 + (u^2 + iu)\eta^2}, \quad G = \frac{\kappa - \rho\eta i u - D}{\kappa - \rho\eta i u + D}, \quad \beta_{t^*} = \frac{\eta^2}{4\kappa}[1 - e^{-\kappa t^*}]$$

Proof. Let's consider the forward log-return $x_{t^*,T} := x_T - x_{t^*} = \log(\frac{S_T}{S_{t^*}})$, the forward starting option price is given by:

$$V_{FS}(t_0) := \mathbb{E}^{\mathbb{Q}}[e^{-rT}(e^{x_{t^*,T}} - e^k)^+ | x_0, v_0]$$
under the $\mathbb{Q}$ density
$$q_{t^*,T}(x)dx := \mathbb{Q}(x_{t^*,T} \in [x, x+dx), | x_0, v_0),$$

By the same notation used in section A.1.1, the exponential dampened option price can be written as

$$v_{FS}(t_0) := e^{\alpha k} \cdot V_{FS}(t_0) = e^{\alpha k}\int_{\infty}^{\infty} e^{-rT}(e^{x_{t^*,T}} - e^k)^+ q_{t^*,T}(x)dx, \quad \text{for some} \quad \alpha > 0$$

Comparing the section A.1.1, we derive the fourier transform of option price following the idea of Carr-Madan. The difference between this expression for forward-starts and expression for normal option price are two folds. One the one hand, the free variable $x_{t^*,T}$ denotes for the difference between the log-price at two ends; on the other hand, the density $q_{t^*,T}(x)$ here is a conditional density conditioning on the initial variance at t_0. To ease the notation, we denote x for $x_{t^*,T}$. Taking the fourier transform of the exponential dampened option price, we obtain the following,

$$\begin{aligned}
\psi_{t^*,T}(u) &:= \int_{-\infty}^{\infty} e^{iuk} \int_{k}^{\infty} e^{\alpha k} e^{-rT}(e^x - e^k) q_{t^*,T}(x) dx dk \\
&= \int_{-\infty}^{\infty} e^{-rT} q_{t^*,T}(x) \int_{-\infty}^{x} (e^{x+\alpha k} - e^{(1+\alpha)k}) e^{iuk} dk dx \\
&= \int_{-\infty}^{\infty} e^{-rT} q_{t^*,T}(x) [\frac{e^{(\alpha+1+iu)x}}{\alpha + iu} - \frac{e^{(\alpha+1+iu)x}}{\alpha + 1 + iu}] dx \\
&= \frac{e^{-rT}\phi_{t^*,T}(u - (\alpha+1)i)}{\alpha^2 + \alpha - u^2 + i(2\alpha+1)u}
\end{aligned}$$

Here is the fourier transformed price in terms of the conditional characteristic function $\phi_{t^*,T}(u)$ of the underlying (diffusion) process. Subsequently, the conditional or forward characteristic function is computed as follows:

$$\begin{aligned}
\phi_{t^*,T}(u) &= \mathbb{E}^{\mathbb{Q}}[\exp(iu(x_T - x_{t^*}))|x_0, v_0] \\
&= \mathbb{E}^{\mathbb{Q}}\Bigg[\underbrace{\mathbb{E}^{\mathbb{Q}}[\exp(iu(x_T - x_{t^*}))|x_{t^*}, v_{t^*}]}_{\phi(u) * e^{-iux_{t^*}}} |x_0, v_0\Bigg] \\
&= \mathbb{E}^{\mathbb{Q}}\Bigg[\exp\{iur\tau + A(u,\tau) + B(u,\tau)v_{t^*}\}|x_0, v_0\Bigg] \\
&= e^{iur\tau + A(u,\tau)} \cdot \underbrace{\mathbb{E}^{\mathbb{Q}}\Bigg[e^{B(u,\tau)v_{t^*}}|v_0\Bigg]}_{\text{laplace transform}}
\end{aligned}$$

Where we denote τ as $T - t^*$. The $\mathbb{C}$ valued (complex) functions $A(u,\tau), B(u,\tau)$ are defined exactly the same as in the Lemma A.1.1. As we can see from the last line, if we can solve the laplace transform to integrate out the v_{t^*} term, we obtain the forward conditional characteristic function. This laplace transform is performed over the square root process [1]and it can be solved by the help of the following lemma.

LEMMA C.2.2 (Laplace transform of transitional PDF of CIR process (Shreve 2005)[92])**.** *For X_t given by*

$$dX_t = (a - bX_t)dt + \eta\sqrt{X_t}dW_t$$

[1]This is called the CIR process and mean reverting process in the financial literature, and can also be defined as a sum of squared Ornstein-Uhlenbeck processes. The process value at time t, i.e. v_t follows a noncentral chi-square distribution

It has distribution of non-central Chi-squared distribution with R degrees of freedom and non-centrality parameter Λ. And its moment generating function (the laplace transformation) is in closed form such that

$$\phi(u) = \mathbb{E}^{\mathbb{Q}}[e^{uX_t}|X_0] = \left(\frac{1}{1-2\beta_t u}\right)^{\frac{2a}{\eta^2}} \exp\left[\frac{e^{bt}uX_0}{1-2\beta_t u}\right]. \text{ for all } u < \frac{1}{2\beta_t} \tag{C.3}$$

where $\beta_t = \frac{\eta^2}{4b}[1-e^{-bt}]$

In the Heston model, $a = \kappa\theta, b = \kappa$, thus we get the laplace transform of the variance process:

$$\mathbb{E}^{\mathbb{Q}}\left[e^{B(u,\tau)v_{t^*}}|v_0\right] = \left(\frac{1}{1-2\beta_t B(u,\tau)}\right)^{\frac{2\kappa\theta}{\eta^2}} \exp\left[\frac{e^{\kappa t}B(u,\tau)v_0}{1-2\beta_t B(u,\tau)}\right]. \tag{C.4}$$

where $\beta_t = \frac{\eta^2}{4\kappa}[1-e^{-\kappa t}]$ Putting pieces together we get the forward (conditional) characteristic function under the Heston model and prove the lemma. □

Appendix D

Determine Weights for Calibration Procedure

In this chapter we are going to discuss about the issue of choosing proper weights for calibration. The criterion considered here is that we should choose the optimization weights such that the difference between the evolution of the option prices suggested by the model and the option prices in the market is the smallest. The approach taken here is to replicate payoff into trackable pieces and then apply Itô formula to track the dynamics of each component.

D.1 Model and Price Dynamics

DEFINITION D.1.1 (Price of a derivative)**.** *The price of a derivative is the total cost of an hedging portfolio that replicates the derivative payoff. We denote such a quantity as* $H(t, S_t, v_t)$.

In the same vein as the portfolio replication argument in [52], we construct a hedging portfolio to replicate the option payoff,

$$H(t, S_t, v_t) = (H_0 + \int_0^t \Delta_t dS_t + \int_0^t \Delta_1 dv_t),$$

(H_0 : martingale component in portfolio, e.g. bond or residual cash)

Specify a general stochastic volatility model, s.t.

$$dS_t = S_t\sqrt{v_t}d(\rho W_t^1 + \sqrt{1-\rho^2}W_t^2)$$
$$dv_t = a(v_t)dt + b(v_t)dW_t^1$$

Note that by choosing $a(v_t) = \kappa(\theta - v_t)$ and $b(v_t) = \eta\sqrt{v_t}$ we get the Heston model.
Then apply Itô's lemma to the portfolio value process $H(t, S_t := s, v_t := z)$, we get

$$dH(t,s,z) = \underbrace{\partial_z H(t,s,z)bdW_t^1 + \partial_s H(t,s,z)s\sqrt{z}d(\hat{\rho}W_t^1 + \sqrt{1-\hat{\rho}^2}W_t^2)}_{\text{Martingale component } dM_t} \tag{D.1}$$
$$+ \partial_t H(t,s,z)dt + \partial_z H(t,s,z)a(v_t)dt + \frac{1}{2}\partial_{ss}H(t,s,z)s^2 z dt \tag{D.2}$$
$$+ \frac{1}{2}\partial_{zz}H(t,s,z)b^2(v_t)dt + \partial_{sz}H(t,s,z)b\rho s\sqrt{z}bdt \tag{D.3}$$

For pricing purpose, we assume that the price process of such portfolio has the same value process as a numéraire asset under risk neutral measure Q. However, in general, this hedging portfolio can not hedge the expected payoff perfectly. Because the hedge ratio depends on the model specified, and models are imperfect because of the estimation error and model error.
So we denote such error as Γ_t and write the dynamics of the real value of such portfolio $\hat{H}(t, S_t, v_t)$ as:

$$\hat{H}(t,S_t,v_t) = \underbrace{(H_0 + \int_0^t \Delta_t dS_t + \int_0^t \Delta_1 dv_t)}_{\text{Expected Price in the Heston model } H(t,S_t,v_t)} + \Gamma_t \tag{D.4}$$

If the market behaves like Heston dynamics with specified parameter χ, then the $\Gamma_t = 0$. Apparently, the Γ_t introduces the 'swing in the profit/loss in the value computation of a portfolio' [20].

D.2 The Effects of Wrongly Specified Stochastic Volatility Model

Here we denote an unknown true stock process $\hat{S}_t$ under risk neutral measure and an unknown true variance process $\hat{v}_t$ with hypothetical correct parameter specification, such

that

$$d\hat{S}_t = \hat{S}_t\sqrt{\hat{v}_t}d(\hat{\rho}W_t^1 + \sqrt{1-\hat{\rho}^2}W_t^2) \tag{D.5}$$

$$d\hat{v}_t = \hat{a}_t(\hat{v}_t)dt + \hat{b}_t(\hat{v}_t)dW_t^1 \tag{D.6}$$

Here we assume that the $\hat{a}_t(v_t)$ and $\hat{b}_t(v_t)$ are non-affine state-dependent (in contrast to affine state-dependency in the Heston model), and the ρ_t is assumed to be deterministic for simplicity. For a general derivative payoff $\Phi(\hat{S}_T) : R^+ \to R^+$, we have the value (or price) process as

$$\hat{H}(t,s,z) = \mathbb{E}[\Phi(\hat{S}_T)|\hat{S}_t = s, \hat{v}_t = z]$$

. Another interpretation of such specification of the real price dynamics are the real parameters are uncertain and follows a process on their own.

Since it is an expected payoff of a tradable asset, so it is a local martingale under risk neutral measure, such that by applying the Feynmann-Kac theorem, the following holds

$$0 = \partial_t\hat{H}(t,s,z) + \partial_z\hat{H}(t,s,z)\hat{a}_t(z) + \frac{1}{2}\partial_{ss}^2\hat{H}(t,s,z)s^2z \tag{D.7}$$

$$+\frac{1}{2}\partial_{zz}^2\hat{H}(t,s,z)\hat{b}_t^2(z) + \partial_{sz}^2\hat{H}(t,s,z)\hat{\rho}_t s\sqrt{z}\hat{b}_t(z) \tag{D.8}$$

So that the *infinitesimal generator* of the real value dynamics of such portfolio is

$$\partial_t\hat{H}(t,s,z) = -\mathcal{A}\hat{H}(t,s,z) \tag{D.9}$$

$$= -\left[\hat{a}_t(z)\partial_z + \frac{1}{2}s^2z\partial_{ss}^2\right. \tag{D.10}$$

$$\left. + \hat{b}_t^2(z)\frac{1}{2}\partial_{zz}^2 + \hat{\rho}_t s\sqrt{z}\hat{b}_t(z)\partial_{sz}^2\right]\hat{H}(t,s,z) \tag{D.11}$$

Now let's assume, *we nearly perfect fitted the model to the market at time τ and the quantity S_τ, v_τ and the derivative prices $\hat{H}(\tau, S_\tau, v_\tau)$ are all observable.* So that at this time point the model reprices the current market states, s.t. $H(\tau, S_\tau, v_\tau) = \hat{H}(\tau, S_\tau, v_\tau)$. Then by equation D.4, we know that from the current time point τ, the real price process will be different from what the model predicts by a quantity of Γ_t. The way of quantifying such differences is let the $\partial_t\hat{H}(t,s,z)$ evolves in a way that is consistent with the infinitesimal generator of the real portfolio value process $\partial_t\hat{H}(t,s,z) = -\mathcal{A}\hat{H}(t,s,z)$.

Therefore, substitute this identity into equation D.1, we have

$$\begin{aligned} d\hat{H}(\tau, S_\tau, v_\tau) =& dM_\tau \\ &+ \partial_z \hat{H}(\tau, S_\tau, v_\tau)(a(v_\tau) - \hat{a}_\tau(v_\tau))dt \\ &+ \frac{1}{2}\partial_{zz}\hat{H}(\tau, S_\tau, v_\tau)(b^2(v_\tau) - \hat{b}^2_\tau(v_\tau))dt \\ &+ \partial_{sz}\hat{H}(\tau, S_\tau, v_\tau)S_\tau\sqrt{v_\tau}(\rho b - \hat{\rho}\hat{b}_\tau(v_\tau))dt \end{aligned}$$

The last part is the difference between model implied value dynamics to the real price process in an infinitesimal time, $d\Gamma_\tau$.

The calibrated parameters a, b, ρ can be seen as a map from the observable quantity to the parameters, such that

$$a(v_t), b(v_t), \rho : C_1^{mkt}(\xi) \times C_2^{mkt}(\xi) \times \dots C_n^{mkt}(\xi) \to \mathbb{R}$$

(C_i^{mkt} denotes the market option prices). The practical insight of such formula is that in order to make the difference Γ_τ small, we have to make sure that the calibrated quantity close to the true quantity in the region where the partial derivative of the $\hat{H}$ w.r.t $\partial_z, \partial_{zz}, \partial_{sz}$ is large. It suggests that we can use the partial derivative as the weights for the calibration procedure.

D.2.1 Numerical Evaluation of the Weights

As we know from the Carr-Madan inversion scheme [23], the European payoffs can be calculated by the knowledge of characteristic function, $\phi_T(u)$, such that

$$\begin{aligned} C_T(k) &= \frac{\exp(-\alpha k)}{\pi}\int_0^\infty e^{-iuk}\psi_T(u)du \\ &= \frac{\exp(-\alpha k)}{\pi}\int_0^\infty e^{-iuk}\frac{e^{-rT}\phi_T(u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+i(2\alpha+1)u}du \end{aligned}$$

Taking derivative over European payoffs can be calculated by taking derivative over characteristic function w.r.t $\partial_z, \partial_{zz}, \partial_{sz}$ first, than put in Carr-Madan inversion scheme.

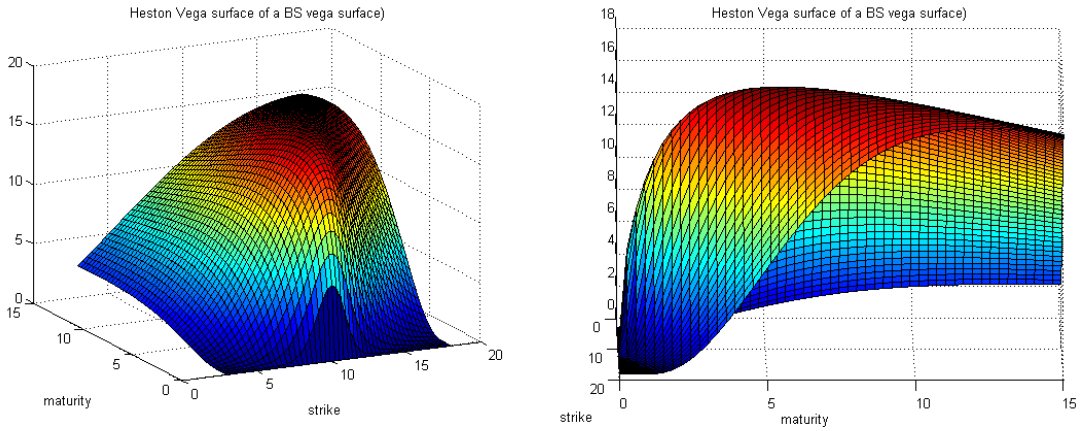

Figure D.1: Heston generated European Call Option Vega surface is of a Black-Scholes Vega shape.

Taking partial derivative over the characteristic function yields the following expressions:

$$\partial_z \phi_T(u; z := v_0) = \exp\left(\frac{1}{\eta^2}\left[\frac{1-e^{-DT}}{1-Ge^{-DT}}\right](\kappa - \rho\eta i u - D)\right) \times \phi_T(u; z := v_0)$$
$$\partial_{zz} \phi_T(u; z := v_0) = \exp^2\left(\frac{1}{\eta^2}\left[\frac{1-e^{-DT}}{1-Ge^{-DT}}\right](\kappa - \rho\eta i u - D)\right) \times \phi_T(u; z := v_0)$$
$$\partial_{sz} \phi_T(u; z := v_0) = iu \exp\left(\frac{1}{\eta^2}\left[\frac{1-e^{-DT}}{1-Ge^{-DT}}\right](\kappa - \rho\eta i u - D)\right) \times \phi_T(u; z := v_0)$$

Applying the Carr-Madan inversion scheme, we can obtain the weights for options of different maturities and strikes:

$$\begin{aligned}\partial_z C_T(k) &= \frac{\exp(-\alpha k)}{\pi}\int_0^\infty e^{-iuk}\partial_z \psi_T(u) du \\ &= \frac{\exp(-\alpha k)}{\pi}\int_0^\infty e^{-iuk}\frac{e^{-rT}\partial_z \phi_T(u-(\alpha+1)i)}{\alpha^2+\alpha-u^2+i(2\alpha+1)u} du\end{aligned}$$

If we omit the higher order terms, like $\partial_{zz}, \partial_{sz}$, but keep only the ∂_z term, we arrived at the same conclusion as Cont (2005) [28], that the weight should be chosen as Vega in order to make the deviation of model implied price evolution to the real price dynamics the smallest.

Bibliography

[1] Albanese C., Campolieti G., 2006, *Advanced Derivatives Pricing and Risk Management*, ELSEVIER Academic Press, London.

[2] Aliprantis C.D., Border K.C., 1999, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Second Edition, Springer.

[3] Andersen L.B.G., Andreasen J., 2002, *Volatile Volatilities*, Risk, 15(12):163-168, December 2002.

[4] Atchade Y.F., J.S. LIU, 2004, *The Wang-Landau Algorithm for Monte Carlo computation in general state spaces*, Technical Report, Harvard University.

[5] Avellaneda, M., Buff, R., Friedman, C., Grandchamp, N., Kruk, L., Newman, J., 2000, *Weighted Monte Carlo: A new Technique for Calibrating Asset-Pricing Models*, IJTAF 2000, Proc. Courant Seminar Vol II.

[6] Bakshi, G., Cao, C. & Chen, Z., 1997, *Empirical performance of alternative option pricing models*, Journal of Finance 52, 20032049.

[7] Bakshi, G., & Chen, Z., 1997, *An alternative valuation model for contingent claims*, the journal of financial Economics.

[8] Bakshi, G., Kapadia, N., 2003, *Delta-Hedged Gains and the Negative Volatility Risk Premium*, Review of Financial Studies 16(2), 527-566.

[9] Bates, D., 1996, *Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark options*, Reviews of Financial Studies, 9: 69-107.

[10] Berg, B.A., Neuhaus, T., 1992, *Multicanonical ensemble: A new approach to simulate first-order phase transitions*, Phys. Rev. Lett. 68.

[11] Black, F., Scholes, M., 1973, *The pricing of options and Corporate Liabilities*, Journal of Political Economy, 81, pp. 637-59, 1973.

[12] Booker, A.J., Dennis, J.E., Frank,P.D., 1998, *Optimization using Surrogate Objectives on a Helicopter Test Example*, Computational Methods in Optimal Design and Control, pp 49-58, 1998.

[13] Borak, S., Detlefsen, K., & Härdle, W., 2005, *FFT Based Option Pricing*, HU Berlin discussion paper.

[14] Boltzmann distribution, http://en.wikipedia.org/wiki/Boltzmann_distribution.

[15] Breeden, D., Litzenberger, R., 1978, *Prices of state-contingent claims implicit in option prices*, Journal of Business, 51, 621-651.

[16] Buchen, P.W. and Kelly, M., 1996, *The maximum entropy distribution of an asset inferred from option prices*, Journal of Financial and Quantitative Analysis, 31 (1).

[17] Buehler, H., 2005, *Expensive Martingales*, Quantitative Finance, Volume 6, Number 3/June 2006 (first version March 2004).

[18] Buehler, H., 2005, *Valuing and Hedging Equity Derivatives*, available at http://www.math.tu-berlin.de/ buehler/.

[19] Buehler, H., 2006, *Consistent variance curve models. Finance and Stochastics*, 10:178-203, 2006.

[20] Buehler, H., 2006, *Volatility Markets: Consistent Modeling, Hedging and Practical Implementation,* , Ph.D. Dissertation, TU Berlin & GME Quantitative Products Analytics, Deutsche Bank, London.

[21] Buehler, H., 2007, *Hedging Options On Variance*, Global Derivatives & Risk Management, Paris, May 22nd, 2007.

[22] Carr, P., Madan, D.B., 1998, *The variance gamma process and option pricing.* European Financial Review 2:79-105.

[23] Carr, P., Madan, D.B., 1999, *Option Valuation Using the Fast Fourier Transform*, journal of Computational finance 2, 61-73.

[24] Chance, D.M., 2007, *An introduction to Derivatives and Risk Management*, Mason Ohio: Thomson South-Western.

[25] Cheyette, O., 1996, *Implied prepayments: addressing anomalies in the modeling of MBS value and risk*, The journal of Portfolio management 23(1).

[26] Cochrane, J.H. and Saa-Requejo, J., 2000, *Beyond Arbitrage: Good Deal Asset Price Bounds in Incomplete Markets*, Journal of Political Economy, 108, 79-119.

[27] Cont, R., Tankov, P., 2006, *Retrieving Levy processes from option prices: Regularization of an ill-posed inverse problem*, SIAM Journal on Control and Optimization.

[28] Cont,R., 2005, *Recovering volatility from option prices by evolutionary optimization*, Research Paper Series.

[29] Cont, R., Tankov, P., Voltchkova, E., 2005, *Hedging with options in models with jumps*, Abel symposium 2005.

[30] Cox, J., Ross, S., 1976, *The valuation of options for alternative stochastic processes.* Journal of Financial Economics, 3:145-166, 1976.

[31] Cox, J. C., Ingersoll, J. E. & Ross, S. A., 1985, *A theory of the term structure of interest rates*, Econometrica 53, 385¨C407.

[32] Cover, J. & Thomas, J.A., 1991, *Elements of Information Theory*, New York, John Wiley & Sons.

[33] Dayal, P., Trebst, S., & Wessel, S., Würtz, D., 2004, *Performance Limitations of Flat Histogram Methods* Phys. Rev. Lett. 92, 097201.

[34] Deutsche Bank, Global Equities Global Quantitative Research, 2002, *Pricing and Hedging under Stochastic Volatility Models*, ppt. Risk 2002, 23rd-24th April, Paris.

[35] Derman, E., 2006, *Modeling the Volatility Smile*, Stanford Financial Seminars 2006.

[36] Duffie, D., Pan, J., and Singleton, K., 2000, *Transform analysis and asset pricing for affine jump diffusions.* Econometrica 68, 1343-1376.

[37] Dupire, B., 1994, *Pricing with a smile*, Risk, 7:18-20.

[38] Engle, R., 1982, *Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation*, Econometrica, 50, 987-1008.

[39] Fletcher, R., *Practical Methods of Optimization*, Wiley, New York, 1987.

[40] Follmer, H., and Sondermann, D., 1986, *Hedging of Non-Redundant Contingent Claims*, Contributions to Mathematical Economics, 205-223.

[41] Fogel, D. B., 2000, *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence (2nd ed.*, IEEE Press, Piscataway, NJ. Wiley, New York.

[42] Follmer, H., and Leukert, P., 2000, *Efficient Hedging: Cost versus Shortfall Risk*, Finance and Stochastic, 4, 117-146.

[43] Gatheral, J., 2004, *Lecture 1: Stochastic volatility and local volatility*, Case Studies in Financial Modeling Notes, Courant Institute of Mathematical Sciences.

[44] Gatheral, J., 2006, *The Volatility Surface: a Practitioner's Guide*, WILEY Finance Series.

[45] Geman, S. and Geman, D. 1984, *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721-741.

[46] Geyer, C.J., *Markov chain monte carlo maximum likelihood*, Computing Science and Statistics: The 23rd symposium on the interface, Interface Foundation, Fairfax, pp.156-163.

[47] Geyer, C. and Thompson, E., 1995, *Annealing Markov Chain Monte Carlo with applications to ancestral inference*, Journal of the American Statistical Association 90: 909-920.

[48] Gilmore, P., Kelley, C. T., *An Implicit Filtering Algorithm for Optimization of Functions with Many Local Minima*, SIAM Journal of Optimization, 5:269-285, 1995.

[49] Grefenstette,J.J., 1992, *Deception considered harmful*, In Foundations of Genetic Algorithms 2, Morgan Kaufmann, San Mateo.

[50] Heath, D., Schweizer, M., 2000, *Martingale versus PDEs in Finance: An Equivalence Result with Examples*, Journal of Applied Probability 37 (2000), 947-957.

[51] Heilmann, F., 2005, *The State Space of Complex Systems*, Ph.D thesis, Technische Universit ät Chemnitz, Fakultät für Naturwissenschaften.

[52] Heston, S.L., 1993, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The review of Financial studies 6(2), 327-343.

[53] Hobson, D.G., Rogers, L.C.G., 1998, *Complete Models with Stochastic Volatility*, Math Finance, 8, 193-205,

[54] Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor.

[55] Hong, G., 2004, *Forward Smile and Derivative Pricing*, Equity Quantitative Strategists Working Paper, UBS.

[56] Hukushima, K., Nemoto, K., 1996, *Exchange Monte Carlo method and application to spin glass simulations*, Journal of Physics Society Japan, 65, 1604-1608.

[57] Hull, J.C., 2005, *Options, Futures and Other Derivatives*(6th edition), Prentice Hall.

[58] Jacquier, A., Slaoui, S., 2007, *Variance Swap and Correlation Swaps*, Available at SSRN: http://ssrn.com/abstract=998924.

[59] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P., 1983, *Optimization by simulated annealing*, Science 220: 671-680.

[60] Lai, T.L. 2003, *Stochastic Approximation*, The annals of Statistics, 31, 391-406.

[61] Lee, R., 2004, *Option pricing by transform methods: extensions, unification and error control*, Journal of Computational Finance, 7.

[62] Paninski,L., *Markov Chain Monte Carlo and Gibbs Sampling*, Lecture Notes for EEB 581, Statistics department at Columbia University.

[63] Liang, F. and Wong, W. H., 2000, *Evolutionary Monte Carlo: Application to c_p model sampling and change point problem*, Statistica Sinica 10(2): 317-342.

[64] Liang, F., 2003, *Annealing contour monte carlo for structure optimization in an off-lattice protein model*, Technical Report .

[65] Liang, F., Liu, C. and Carroll, R.J., 2007, *Stochastic Approximation in Monte Carlo Computation*, J. Amer. Statist. Assoc., 102, 305-320.

[66] Lindström, E., Ströjby, J., Brodén, M., Wiktorsson, M., 2007,*Sequantial Calibration of Options*, Lund Unversity Working Paper.

[67] Liu, J.S., Liang, F. & Wong, W.H., 2000, *The Use of Multiple-Try Method and Local Optimization in Metropolis Sampling*, J. Amer. Statist. Assoc. Vol. 95, No.449, 121-134.

[68] Liu, J.S., Sabatti, C., 2000, *Generalised Gibbs sampler and multigrid Monte Carlo for Bayesian computation*, Biometrika, 2000, http://citeseer.ist.psu.edu/liu99generalized.html.

[69] Liu, , *Monte Carlo Strategies in Scientific Computing*, Springer Series in Statistics.

[70] Liu, J.S., Wong, W., and Kong, A., 1994, *Covariance Structure of the Gibbs Sampler with Applications to the Comparisons of Estimators and Augmentation Schemes*, submitted to Biometrika.

[71] Liu, J.S., 1994, *The collapsed Gibbs sampler with applications to a gene regulation problem*, Journal of the American Statistical Association 89 958-966.

[72] Merton, R., 1976, *Option Pricing When the underlying Stock Returns are Discontinuous*, Journal of Financial Economics 4, 125-144.

[73] MacKay, D., 2003, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press.

[74] Ng, M.W., 2005, *Option Pricing via the FFT*, Master Thesis, Applied Institute of Mathematics, TU Delft.

[75] Merton, R., 1973, *The theory of rational option pricing*, Bell J. Econ. Manag. Sci.4, pp.141-183.

[76] Powell, M.S.D., 1988, *A Review of Algorithms for Nonlinear Equations and unconstrained Optimization*, pp.220-232 in ICIAM 1987 Proceedings, SIAM, Philadelphia 1988.

[77] Rebonato, R., 2003, *Term-Structure Models: a Review*, Royal Bank of Scotland Quantitative Research Centre Working papers.

[78] Rebonato, R., 2004, *Volatility and Correlation, 2nd Edition*, John Wiley & Sons, New York.

[79] Roberts, G.O., Tweedie, R.L., 1995, *Exponential Convergence of Langevin Diffusions and Their Discrete Approximations*, Bernoulli 2 341-362.

[80] Roberts, G.O. & Sahu, S.K., 1996, *Updating Schemes, Correlation Strucure, Blocking and Parameterization for the Gibbs Sampler*, Working paper of University of Cambridge.

[81] Robert, C.P., Casella, G., 1999, *Monte Carlo statistical methods*, Springer-Verlag, New York.

[82] Roberts, G.O., Rosenthal, J.S., *Optimal scaling for various Metropolis-Hastings algorithms*, Statist. Sci. Volume 16, Number 4 (2001), 351-367.

[83] Robbins, H. and Monro, S. 1951, *A Stochastic Approximation Method*, Annals of Mathematical Statistics, 42, 501-514.

[84] Samperi, D.J., 1997, *Inverse Problems, Model Selection, and Entropy in Derivative Security Pricing*, PhD Thesis, Department of Mathematics, NYU.

[85] Shoutens, W., Simons, E., and Tistaer, J., *A Perfect Calibration! Now what? Model risk for exotic and moment derivatives*, Frankfurt Financial Seminar 2005.

[86] Scott, L., 1997, *Pricing Stock Options in a Jump-Diffusion Model with Stochastic Volatility and Interest Rates: Application of Fourier Inversion Methods*, Mathematical Finance.

[87] Shell, M.S., Debenetti, P.G., Panagiotopoulos, A.Z., 2002, *Generalization of the wang-landau method for off-lattice simulations*, Physical review E 66.

[88] Shell, M.S., Debenetti, P.G., Panagiotopoulos, A.Z., 2003, *An improved monte carlo method for direct calculation of the density of states*, Technical Report.

[89] Shephard, N., Pitt, M.K., 1997, *Likelihood analysis of non-Gaussian measurement time series*, Biometrika 84, 653-67.

[90] Shleifer, A., 2000, *Inefficient Markets – An Introduction to Behavioural Finance*, Clarendon Lectures in Economics, Oxford University Press, Oxford.

[91] Karatzas, I., & Shreve, S.E., 1988, *Brownian Motion and Stochastic Calculus (2d ed.)*, Springer-Verlag, New York.

[92] Shreve, S.E., 2005, *Stochastic Calculus Models for Finance: Continuous Time Models*, Springer Finance Series, Springer New York.

[93] Spall, J.C., 2001, *Introduction to Stochastic Search and Optimization*, Wiley-Interscience Series in Discrete Math and optimization.

[94] Steele, M., 2001, *Stochastic Calculus and Financial Application*, Application of Mathematics series, Springer.

[95] Taleb, N., 1996, *Dynamic Hedging: Managing Vanilla and Exotic Options*, John Wiley & Son, New York.

[96] Wang, F. and Landau, D.P., 2001, *Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States*, Physical Review Letters, 86, 2050-2053.

[97] Wilmott, P., 2002, *Cliquet Options and Volatility Models*, Wilmott Magazine, December, 2002.

[98] Wolpert, D.H., Macready, W.G., 1997, *No Free Lunch Theorems for Optimization*, IEEE Transactions on Evolutionary Computation, vol. 1, pp. 67-82.

[99] Zhou, C., Bhatt, R.N., 2003, *Understanding and improving the wang-landau algorithm*, Technical Report.

[100] Zhu, C., Boyd, R.H., Lu, P., and Nocedal, J., 1994, *L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization*, Northwestern University, Department of Electrical Engineering.