UNCERTAINTY QUANTIFICATION OF PARAMETERS IN LOCAL VOLATILITY MODEL VIA FREQUENTIST, BAYESIAN AND STOCHASTIC GALERKIN METHODS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ABDULWAHAB ANIMOKU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
FINANCIAL MATHEMATICS

SEPTEMBER 2018

Approval of the thesis:

UNCERTAINTY QUANTIFICATION OF PARAMETERS IN LOCAL VOLATILITY MODEL VIA FREQUENTIST, BAYESIAN AND STOCHASTIC GALERKIN METHODS

submitted by **ABDULWAHAB ANIMOKU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Financial Mathematics Department, Middle East Technical University** by,

Prof. Dr. Ömür Uğur
Director, Graduate School of **Applied Mathematics** ⎯⎯⎯⎯⎯⎯

Prof. Dr. Sevtap Selçuk-Kestel
Head of Department, **Financial Mathematics** ⎯⎯⎯⎯⎯⎯

Prof. Dr. Ömür Uğur
Supervisor, **Scientific Computing, METU** ⎯⎯⎯⎯⎯⎯

Examining Committee Members:

Prof. Dr. Sevtap Selçuk-Kestel
Actuarial Science, METU ⎯⎯⎯⎯⎯⎯

Prof. Dr. Ömür Uğur
Scientific Computing, METU ⎯⎯⎯⎯⎯⎯

Assoc. Prof. Dr. Ümit Aksoy
Mathematics, Atılım University ⎯⎯⎯⎯⎯⎯

Assist. Prof. Dr. Hamdullah Yücel
Scientific Computing, METU ⎯⎯⎯⎯⎯⎯

Assist. Prof. Dr. Ahmet Şensoy
Business Administration, Bilkent University ⎯⎯⎯⎯⎯⎯

Date: ⎯⎯⎯⎯⎯⎯

iv

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ABDULWAHAB ANIMOKU

Signature   :

ABSTRACT

UNCERTAINTY QUANTIFICATION OF PARAMETERS IN LOCAL VOLATILITY MODEL VIA FREQUENTIST, BAYESIAN AND STOCHASTIC GALERKIN METHODS

Animoku, Abdulwahab

Ph.D., Department of Financial Mathematics

Supervisor   : Prof. Dr. Ömür Uğur

September 2018, 134 pages

In this thesis, we investigate and implement advanced methods to quantify uncertain parameter(s) in Dupire local volatility equation. The advanced methods investigated are Bayesian and stochastic Galerkin methods. These advanced techniques implore different ideas in estimating the unknown parameters in PDEs. The Bayesian approach assumes the parameter is a random variable to be sampled from its posterior distribution. The posterior distribution of the parameter is constructed via "Bayes theorem of inverse problem". Stochastic Galerkin method involves propagating uncertainty into a deterministic input parameter and then quantifying the randomness in the solution. In addition, the performance and numerical analysis of each approach are studied.

Keywords: Local volatility, Bayesian analysis, Stochastic Galerkin method, Tikhonov regularization

ÖZ

YEREL OYNAKLIK MODELİ PARAMETRELERİNİN SIKLIKÇI, BAYESÇİ, VE STOKASTIK GALERKİN YÖNTEMLERİ İLE BERİRSİZLİK ÖLÇÜMÜ

Animoku, Abdulwahab

Doktora, Finansal Matematik Bölümü

Tez Yöneticisi   : Prof. Dr. Ömür Uğur

Eylül 2018, 134 sayfa

Bu tezde Dupire lokal oynaklık denklemindeki belirsiz parametreleri ölçmek için kullanılan gelişmiş teknikleri inceledik ve uyguladık. Araştırılan ileri yöntemler Bayes' ve stokastik Galerkin yöntemleridir. Bu gelişmiş teknikler, kısmi differansiyel denklemlerin bilinmeyen parametrelerini tahmin etmek için farklı yöntemler kullanırlar. Bayes' yaklaşımı parametrenin sonsal (posterior) dağılımından örneklenecek rastgele bir değişken olduğunu varsayar. Parametrenin sonsal dağılımı "ters problemin Bayes' teoremi" ile oluşturulur. Stokastik Galerkin yöntemi, belirsizliği deterministik bir girdi parametresine yaymayı ve sonra çözümdeki rasgeleliği ölçmeyi içerir. Ayrıca her bir yaklaşımın performası ve sayısal analizi üzerinde çalışılmıştır.

Anahtar Kelimeler: Yerel oynaklık, Bayes' analizi, Stokastik Galerkin yöntemi, Tikhonov düzenlemesi

To My Family

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

MCMC	Markov Chain Monte Carlo
OLS	Ordinary Least Squares
PC	Polynomial Chaos
PDE	Partial Differential Equation

LIST OF SYMBOLS

$\mathbb{R}$	Real number
$\mathcal{L}_2$	Hilbert space
σ_{LV}	Local volatility
σ_{BS}	Black Scholes implied volatility
K	Strike price
T	Time to maturity
f_t	Model price at time t
V_t	Observed option price at time t
q	Set of parameters
$\mathcal{Q}$	Set of admissible volatility parameters
H	Loss function
ξ	Set of independent normally distributed random variables

CHAPTER 1

INTRODUCTION

In the last two decades, there have been significant improvements in developing financial instruments for investment and hedging purposes. Due to the importance of these instruments for the stability of the economies of countries around the world, it is imperative to correctly price them to avoid arbitrage opportunities. The most recent, infamous 2008 financial crisis is on a large part due to mismanagement and exploitation of financial instruments for arbitrage opportunities. Therefore, active research are being held in the field of Financial Mathematics to capture the behavior of the underlying market processes of these instruments with more exhaustive and sophisticated models. Typically, an agent chooses a model and calibrate it to the observable market prices as a benchmark to avoid mis-pricing. However, due to uncertainty in the calibration procedure and model type, different agents would obtain varied prices. A small mis-pricing among these agents could lead to other market agents making risk-less profits off this mis-pricing.

Today, one of the most commonly used derivative instruments are options. An option can either be a call or put. A call (put) option gives the buyer the right but not the obligation to buy (sell) an underlying at an agreed upon price. An option of the European type only allows exercising of the option at maturity while the American type option allows exercising anytime until maturity. Option as an instrument is widely used as a calibration and hedging instruments. Pricing options involve some set of input parameters, such as, interest rate, volatility, strike price, and maturity date of the option. Among these input parameters,

only volatility is unobservable in the market. Several models such as Black-Scholes, local volatility, and stochastic volatility models that exist in literature try to capture the uncertainty of the underlying asset's volatility. Within these models exist different assumptions around the volatility that create model uncertainty. For example, Black-Scholes model assumes an asset dynamics where the volatility is taken to be constant. This is the simplest form of all the models taken computational time and model complexity into account. Although in this model, the market is complete, it does not explain the volatility smile inherent in the implied volatility of the market option prices.

On the other side, there are the stochastic and jump diffusion volatility models that introduce new sources of randomness into the asset's dynamics. With this additional stochasticity, these models do not replicate every contingent claim in the market. Therefore, in order for these models to reproduce the market prices accurately to a certain degree, there is a trade-off between incompleteness and complexity with numerical accuracy.

In between these two extreme ends is local volatility that operates under the completeness of the market and reproduces a deterministic volatility that explains the smile effect. Thus, local volatility gives the right balance between model complexity and representation of reality.

Other sophisticated models exist, one of which, for instance, the so called hybrid models that synergize the features of two or more models to create a new one. Examples of these are stochastic-local volatility and stochastic-jump diffusion models [2, 47, 50, 53].

Furthermore, calibration of the above financial models does not accurately reproduce all market prices. This can lead to inconsistencies and mis-pricing discussed earlier. On the flip side, there might not be enough observable prices to determine the calibration parameters uniquely from the model. Also, there is a question of stability for a calibrated model–as small change in observable prices can lead to large change in the calibrated parameter. This leads to parameter uncertainty which can create difficulty in hedging risk if the parameter is not properly quantified.

In the current literature, despite the above calibrated problems, most of the efforts are geared towards finding the single unique parameter that reproduces all the prices in the market. Little work is done in checking for the robustness of the calibrated parameter. The stability of the calibrated parameter is important for two reasons, since a disproportionate change in the parameter for a small change in market price can lead to greater risk for any investment made that depends on such parameter. First, it enables proper risk management and other quantitative measures associated with such investments. Second, it gives confidence on the calibration results based on the data set.

In this thesis, we will concentrate on handling these fairly common calibration problems for local volatility model. We do this in two ways: first, we restructure the inverse problem into Bayesian framework to obtain a whole distribution for the calibrated parameter; second, we optimize for the distribution of the parameter using stochastic Galerkin approach.

The Bayesian approach provides a natural way of assigning probabilities to points in the parameter space by combining a set of prior information with likelihood function. The so called *posterior density* constructed via this method can be used for subsequent pricing and hedging of financial instruments such as exotic options. The Bayesian method is a powerful approach since it allows updating of the model as new prices are observed anytime in the market. This creates consistency for the calibrated parameter through time.

For the stochastic Galerkin approach, we quantify the parameter uncertainty by using polynomial chaos expansion as a natural way to represent the observable prices treated as random variables. In this approach, the coefficients of the orthogonal polynomials are first determined through a discretization scheme. Then, by an error minimization procedure, the model parameters are estimated.

1.1 Motivation and Goals

In the previous section, we have already established the importance of the volatility parameter used in pricing of options and other instruments that are used as

investment and hedging tools. The volatility as a calibrated parameter should also exhibit volatility skewness (or sneer) if model is correctly calibrated. One motivation for choosing local volatility model in our study is that it reproduces volatility surface with skewness. On the other hand, the choice of the numerical techniques in handling the inverse problem associated with characterizing the volatility function is also important. In this thesis, we focus on two advanced numerical techniques, namely: Bayesian analysis and stochastic Galerkin method.

In the last decade, there have been few contributions to the literature regarding the use of Bayesian analysis, especially in application to financial problems [25, 26, 45, 46]. Our work is among the few that highlights the importance of formulating and solving a financial problem, especially one with a nonparametric parameter estimation . More uniquely, we have incorporated the use of Markov chain Monte Carlo (MCMC) in sampling the volatility estimates from the posterior distribution. The advantages of having a distribution for volatility via Bayesian analysis is wide and enormous. For example, having a distribution of parameters with each having assigned probabilities allows for construction of volatility surfaces with different confidence levels. Therefore, an agent can subsequently price the financial instruments like options with the estimates in these surfaces with their corresponding confidence levels. Instead of one true parameter value that could lead to a large loss in wealth, the investors choose from several prices corresponding to multiple parameter values with assigned probabilities and confidence levels. We would explore more on this in Chapter 4.

Another goal of this thesis is to add to a growing literature on uncertainty quantification via spectral methods. Dupire local volatility equation can be reformatted into a constrained (parabolic) partial differential equation (PDE). We then solve this PDE constraint via a stochastic Galerkin approach in an optimization procedure that involves, for instance, the use of *Levenberg-Marquardt* algorithm. Also, not much work has been done in using such an advanced and powerful numerical technique in solving financial problems. Therefore, we will explore some theoretical understanding of this methodology from uncertainty quantification perspective. We will also give some insights in applying the concepts to obtain

the calibrated volatility surface.

This thesis is focused on estimating the volatility via Dupire local volatility model by employing various statistical and advanced numerical techniques. The theoretical frameworks of the model and the numerical techniques used will be discussed in details. Consequently, the distribution of volatility surfaces according to each method used will be obtained. In addition, the implications of the results for financial agents and investors will be analyzed and recommendations will be proposed.

1.2 Literature Review

In this section, firstly, we will discuss the developments of various volatility models in the literature. Secondly, we will give a more elaborate understanding of the literature regarding the numerical techniques used throughout this thesis.

One of the earliest analytical models proposed in finance to price a financial instrument (option) is Black-Scholes model [14]. Since then, the variety of financial models has grown more and more. One of the major reasons for this growth can be attributed to the financial crisis that made investors demand more accurate pricing models. In the context of derivative pricing, volatility models were developed to avoid mis-pricing of the financial instruments. Specifically, in 1994, Dupire [33] published his paper on local volatility model. In his model, the volatility is assumed to be a function of asset value S and time t. This produces a more consistent result with the market *volatility smile*. Therefore, the local volatility model reproduces a more realistic volatility surface when compared to the Black-Scholes model. Furthermore, Derman and Kani [28] also contributed independently to the foundation of local volatility model. More on this will be presented later in Chapter 2.

In addition, more complex volatility models have also been proposed. For example, Hull and White [53] introduced the stochastic volatility model, in which a new source of randomness called volatility of volatility is introduced. Over time, different variations of this model have been introduced into the literature,

for example, Heston and SABR models [50, 47]. The use of stochastic volatility models in explaining the smile effect is detailed in [29, 47].

Another important form of volatility model that has been established in literature with far outreach is the jump diffusion model. In such a model, a Lévy or various forms of Poisson processes are used to model the jump sizes. Several authors have studied on the calibration of jump diffusion processes to the market. The works of Merton [69], Zhou [90], and Hilberink et al. [52] highlight these applications.

Meanwhile, we review some of the existing results in literature on the use of Bayesian approach to estimate parameters in a parametric and non-parametric models. The books by Ralph [75], Fitzpatrick [38], and Gelman et al. [40] have been used as the basis of references throughout this thesis. As mentioned earlier, the Bayesian inference have only gained popularity in the last two decades. One of the earlier work dates back to 2002 by Jacquier et al. [57]. The authors showed that the Bayes' estimators of a stochastic volatility model performed better than moments and quasi-likelihood estimators. In the case of interest rate modeling, Bhar et al. [12] have used dynamic Bayesian approaches to calibrating instantaneous spot interest rates.

More recently, the work of Gupta & Reisinger [45, 46], Darsinos & Satchell [25, 26] show the application of Bayesian method to local volatility modeling. The first paper [45], reformats the Dupire local volatility equation into Bayesian inference by constructing a set of posterior estimators for the calibrated volatility parameter. The consistency of the Bayesian estimates of the volatility surfaces were constructed. In [46], the authors measured the risk uncertainty of the Bayesian estimates and the use of these estimates in hedging. In [25], a joint prior distributions for the asset price S_t and the constant implied volatility σ in Black-Scholes model were formed. Therein, the posterior distribution consists of a joint density for asset price S_t and Black-Scholes European call option prices. The marginal density of the option prices are computed from the joint posterior density. Subsequently, in [26], the authors use this marginal density to predict the call option prices for a day ahead. More details can be found in the articles

6

highlighted so far and the references therein.

On the other hand, we discuss some of the literature that extensively explain the application of spectral methods in solving PDEs. In this thesis, we have made use of the books [65, 75, 87] and the references therein as basic tools for understanding the mathematical concepts behind the stochastic Galerkin method. Furthermore, the articles [21, 88] are very helpful in understanding the numerical application of stochastic Galerkin method. However, to the best of our understanding, this method has not been adopted in solving the Dupire local volatility nor has it been used to construct volatility surfaces for other volatility models. Hence, our work will be one of the first application of stochastic Galerkin method in estimating the local volatility surfaces through an optimization approach. However, other methods such as *Finite Difference method* and *Finite Element method* have been used in solving volatility models and pricing strategies. The books by Mikhailov et. al. [70] and Hilber et. al. [51] highlight the use of these methods respectively.

1.3 Outline

The objective of this thesis is to make practical contribution to the calibration of non-parametric financial problem using Bayesian and stochastic Galerkin approaches.

In Chapter 2, we seek to understand the calibration problem. Firstly, we set up the calibration equation as an inverse problem. Furthermore, we discuss the conditions necessary to have a well-posed inverse problem. Consequently, the regularization techniques in handling the ill-posedness of the inverse problem are analyzed. Secondly, we reformat the Dupire local volatility as an inverse problem. Here, we focus on setting up the assumptions guiding the local volatility model as well as the challenges facing the solution of the model. Finally, we give some literature review on different regularization techniques in handling the ill-posedness of the Dupire local volatility.

Chapter 3 focuses on the frequentist approach to parameter estimation. Firstly,

we show the mathematical derivation of the sampling distribution under linear regression model. Secondly, the sampling distribution under the nonlinear parametric model is explained. Finally, in this chapter, we discuss the numerical implementation of frequentist approach in constructing local volatility surfaces.

In Chapter 4, we focus on understanding the theoretical and practical application of Bayesian analysis. Firstly, we discuss the Bayes' formula and different Bayesian estimators. Also, the advantages and challenges using this method are also analyzed. Secondly, we concentrate on the theoretical framework of establishing the consistency of Bayes' estimators. Thirdly, we discuss Markov Chain Monte Carlo method of sampling estimates from a posterior distribution obtained via Bayesian method. Here, we start with the basics of Markov Chains and Metropolis Hasting Algorithm. We give a holistic view on Markov chains and give some theorems regarding the convergence of MCMC estimates. Moreover in this chapter, we highlight the mathematical tools regarding the application of Bayesian method in deriving posterior densities for local volatility parameter via different statistical models and error distributions. Finally, we give numerical examples and interpret the results obtained. The convergence of the MCMC estimates is analyzed and some discussion is made based on the results.

In Chapter 5, we consider the use of stochastic Galerkin method in solving the local volatility model. Firstly, we give details on expressing random variables using polynomial chaos expansion. We also give some examples in understanding this concept using various distributions. Secondly, we discuss the general approach of solving partial differential equations using stochastic Galerkin approach. Moreover in this chapter, we give the mathematical formulation of solving the local volatility model with the numerical results and volatility surface given. Here, we compare the results obtained via the stochastic Galerkin method to Bayesian as well as the surface constructed via Monte Carlo method.

In the final chapter of the thesis, Chapter 6, we summarize the main results of our findings and make some discussions. We also give insight to possible ways to improve our work and outlook for further studies.

CHAPTER 2

CALIBRATION OF LOCAL VOLATILITY

In this chapter, we set up the local volatility as an inverse problem. Here, we discuss the conditions under which the solution of the local volatility model is well-posed. In addition, Tikhonov regularization used in obtaining smooth volatility surface is explained. Furthermore, we mention the various forms of local volatility equations and their practical usage. Finally, we discuss in greater details some existing literature on different methods of solving the local volatility model.

As explained in the previous chapter, the pricing of financial security (for example options) involves an underlying asset process and some input parameters. One of such input parameters is volatility, which measures the fluctuations in the return of the asset price process. This parameter is also termed as the standard deviation of the asset's returns. The volatility is significant in that it is the only input parameter that can not be measured directly from the market, hence, it has to be estimated via a realistic financial model. In the past couple of decades, many of such models have been proposed in the literature. In this thesis, we focus on the Dupire local volatility model. This model characterizes the volatility as a function of asset price level S_t and time t. Given that the motivation behind this model is to find a realistic way of modeling the volatility skewness that Black Scholes model fails to address, the volatility function can be derived in terms of strike price K and maturity T. Therefore, for a single asset, the volatility surface can be obtained to price variety of options with different strike prices and maturities.

However, in order to characterize the local volatility function non-parametrically, we need to calibrate the local volatility model to a market data of option prices. This procedure is an inverse problem. This is so, since we know analytically how to get to the option prices via the Dupire local volatility model but not from the prices to the model parameters. Hence, the calibration problem is an inverse problem of moving from the market prices to the parameter that fit the data. In addition, since the market data is given up to a *bid-ask spread*, the calibrating parameter should be determined accordingly to avoid arbitrage opportunities.

Most of the calibration methods proposed in the literature use a best fit approximation analysis by running a minimization algorithm over the squared errors constrained by the parameter space. For example, in [1, 3, 63], the authors use the minimization approach to finding the best fit solutions in $\mathcal{L}_2$ space. We explore a probabilistic approach of solving this non-parametric inverse problem through Bayesian framework. This approach allows us to assign probabilities to the volatility values obtained at the discretized points on each volatility surface. Furthermore, we can also compute the surfaces using confidence intervals to characterize the risk of each surface. Thus, the Bayesian approach we have adopted in this thesis no longer needs to find the best model that replicates the prices observed in the market. Rather, we solve an entire class of models by obtaining probable class of volatility surfaces that can replicate the market option data within the bid-ask spread.

This chapter is structured as follows: In Section 2.1 and Section 2.2, we formalize the inverse problem set up and discuss the properties of well-posedness and regularization methods. In Section 2.3, we present the local volatility model in different functional forms. Furthermore, in Section 2.4, the error minimization procedure for solving the Dupire equation will be explained. Finally, in Section 2.5, we give some literature review on different methods of calibrating local volatility function.

2.1 Inverse Problem

In this section, we give the general inverse problem set up for parameter esti-
mation. We also discuss the ill-posedness of inverse problem and conditions for
a well-posed problem. Furthermore, we highlight the regularization techniques
used in solving an ill-posed problem.

Many problems in nature are often modeled as an inverse problem. Many sci-
entists and engineers generally desire to relate some set of observations from
measurements or experiments to some physical parameters q, characterizing a
model. Consider the following equation

$$F(q) = d, \tag{2.1}$$

where d is the set of physical observations, F is a function relating q and d, and
q is the set of parameters. There is a distinction in interpretation of (2.1) for
physicists and engineers as compared to applied mathematicians. For physicists
and engineers, they refer to F as the *forward operator* and q as the *model*. On the
other hand, applied mathematicians refer to $F(q) = d$ as a *mathematical model*
and q as the *model parameters*. For our purpose, we adopt the interpretation of
the applied mathematicians to (2.1).

In practice, d can be a discrete observation or a function of time and/or space.
Furthermore, F can be a function or an operator depending on the characteri-
zation of q and d. In addition, the mathematical model in (2.1) is often modeled
with errors. There are two ways to account for the errors in such model. Ei-
ther the errors arise as a result of un-modeled influences on the observations or
they arise due to numerical round off. Thus, given the set of observations d, we
assume the data consists of "perfect" measurements d_{true} and additional noise
term η,

$$\begin{aligned} d &= F(q_{\text{true}}) + \eta \\ &= d_{\text{true}} + \eta, \end{aligned} \tag{2.2}$$

where d_{true} satisfies (2.1) given q_{true}, assuming that the forward mathematical
model is exact.

The *forward problem* involves finding the values of d given model parameters q. In many cases, computing $F(q)$ involves solving an ODE or PDE. However, we are interested in solving the *inverse problem* of estimating q given d.

For a *discreet inverse problem* or *parameter estimation problem*, we determine a finite number of parameters, p, to define a model. As in many cases, we represent the parameters to be estimated as a set of p element vector. Similarly, if there is a finite set of observations n, then we can represent the data as an n element vector. Therefore, we can represent the statistical model in (2.1) as a system of equation given by

$$F(\mathbf{q}) = \mathbf{d}. \tag{2.3}$$

For more detailed information on this subject, interested readers should see [6, 72, 74]. Here, we emphasize inverse problems with finite dimensional parameter space. However, this can be extended to problems with infinite-dimensional parameter space. We restrict our study to finite-dimensional parameter space for practical purpose. Thus, the local volatility function, σ_{LV} will be assumed to be a finite-dimensional vector $\sigma \subset \mathbb{R}^M$, for $M \in \mathbb{N}$. This approximation is justifiable, since we have to discretise the function in order to numerically approximate.

2.2   The Problem Setup

In this section, we set up the inverse problem corresponding to solving non-parametric financial model. Suppose we have a complete market where every contingent claim, h, can be replicated by a self financing portfolio. Suppose further that the contingent claim h is a function of the asset price, S. Furthermore, suppose S follows a price process $S = (S_t)_{t \geq 0}$ that is a function of time t, a stochastic process $W = (W_t)_{t \geq 0}$, and volatility parameter σ, that is,

$$S_t = S\left(S_0, t, (W_u)_{0 \leq u \leq t}; \sigma\right),$$

where S_0 is the initial value of the price process at time $t = 0$. Consequently, S is an $\mathcal{F}$-adapted process given the filtration $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$, generated by the sigma algebra of the price process.

Therefore, at maturity T of the claim, let the pay-off of the claim be given as $h(S(\sigma))$. For specificity, consider an option contract with payoff function h written on $S(\sigma)$ over the time horizon $[0, T]$. The value of the claim at an arbitrary time, $t \in [0, T]$, with respect to the risk neutral probability measure $\mathbb{Q}$ is given by

$$f_t(\sigma) = \mathbb{E}^{\mathbb{Q}} \left[B(t, T) h(S(\sigma)) \mid \mathcal{F}_t \right],$$

where $B(t, T)$ is a discount factor over the interval $[t, T]$. Depending on the model and its assumptions, $B(t, T)$ can be deterministic or stochastic.

In the context of this thesis, the option prices $f_t(\sigma)$ are obtained via local volatility model. Therefore, the forward problem would involve specifying σ in order to compute the prices.

Theorem 2.1 ([45], Remark 1.2). *Suppose we observe a set of prices $\{V_t^{(i)} : i \in I_t\}$, at time $[0, T]$ with noise $\{e_t^{(i)} : i \in I_t\}$, where I_t is an index set. We can represent the inverse problem as reconstructing σ such that*

$$V_t^{(i)} = f_t^{(i)}(\sigma^*) + e_t^{(i)}, \tag{2.4}$$

where σ^* is regarded as the true parameter that reproduces the market prices from the model.

Remark 2.1. The calibration problem involves finding the value of σ that best reproduces the observed prices $\{V_t^{(i)} : i \in I_t\}$. Note, however that there are several methods in literature that have been proposed to solve this problem [48, 49, 63, 35, 27, 23]. Here, we seek such reconstruction using Bayesian and stochastic Galerkin methods.

In the next section, we specify the conditions that render feasibility and stability of solutions for such inverse problem.

2.2.1 Well-Posedness

Here, we first discuss the Hadamard's criteria (see [36, 74] for example) for a *well posed* problem.

Definition 2.1. We say a mathematical problem is well-posed if it satisfies the following properties: for all admissible data

(a) a solution exists (existence),

(b) the solution is unique (uniqueness),

(c) the solution depends smoothly on the given data (regularity).

Given the above conditions, if a mathematical problem violates any of the above criteria we call it an ill-posed problem. More often, parameter estimation problems are ill-posed. Specifically, reconstructing local volatility function is an ill-posed problem for the following reasons: First, there is not enough option data in the market. This inhibits the construction of a volatility surface that continuously depends on the data; secondly, the observed option prices are over dispersed. As a consequence, small changes in the prices can lead to large errors in the partial derivatives of the local volatility equation. To handle the above challenges, there is need to regularize the solution so that condition (c) is satisfied.

For the local volatility calibration problem to satisfy property (a), we assume that we can find a solution fitting the observable prices within a specific tolerance level. Since option prices are often given within a *bid-ask spread*, we assume the true solution lies within this spread. In this context, the spread can be thought of as the given tolerance level.

A major drawback with the calibration of local volatility as mentioned earlier is the insufficient market prices available for the calibration. This limitation allows for more than one parameter to produce model prices that fit the market data. This type of inverse problem is called *underdetermined inverse problem*. On one hand, choosing the wrong parameter can lead to mis-pricing and incorrect hedging of financial instruments. This can often lead to costly losses for a trading agent. On the flip side, if a unique solution is found and it does not continuously depend on the observed market prices, then a small error in one of the market prices can lead to large error in the calibrated parameter. This will lead to mis-pricing again as well as hedging of other instruments. To resolve these challenges,

care should be taken on the numerical method chosen for the estimation process. We have shown that the Bayesian and stochastic Galerkin methods can correctly estimate the calibrated parameter as illustrated in Chapter 4 and Chapter 5 respectively.

Next, we turn to regularization methods to discuss how to overcome the challenges above to ensure the solution follows properties (b) and (c).

2.2.2   Regularization methods

Regularization involves the process of finding an approximating solution to an ill-posed problem by transforming it to a well-posed problem. The most documented method of *regularization* is the *Tikhonov regularization.* In the literature, a vast documentation of this regularization technique exists, see [84, 72, 6] for example. As a consequence, most ill-posed inverse problems in literature are solved using the Tikhonov regularization. Specifically, most of the literature on solving local volatility model use this mothod or a modified version of it. In Chapter 4 and Chapter 5, we propose new approaches to regularize and solve the same problem.

Now, we give the general mathematical structure to regularizing the solution to an ill-posed problem. Consider the general forward problem in (2.1); we are interested in finding the inverse function F^{-1}. Suppose that we can only observe a noisy approximation d^α for the data d, such that $||d^\alpha - d||_\mathcal{D} \leq \alpha$, $d \in \mathcal{D}$. We are instead interested in solving the equivalent problem $F^{-1}(d^\alpha) = q^\alpha$, $q \in \mathcal{Q}$. Assume further that F^{-1} does not satisfy the Hadamard's conditions (b) and/or (c) listed in Section 2.2.1. A common approach is to introduce a *regularization operator* F_λ^{-1} to replace F^{-1} where $\lambda > 0$ is the so called *regularization parameter* which depends on the data d^α and/or α. The operator and parameter are chosen such that the $d^\alpha \to d$ as $\lambda \to 0$.

This can be restated succinctly as

$$\lambda = \lambda(\alpha, d^\alpha) > 0, \quad F_\lambda^{-1} : \mathcal{D} \to \mathcal{Q}; \quad \lim_{\lambda \to 0} \sup \left\{ ||F_\lambda^{-1}(d^\alpha) - F_\lambda^{-1}(d)|| \right\}_\mathcal{Q} = 0.$$

Since it is not always practical and easy to solve $F_\lambda^{-1}(d^\alpha) = q^\alpha$, we instead resolve to solve a more tractable function $G_\lambda(q^\alpha) = F_\lambda(q^\alpha) - d^\alpha$. Therefore, we find q^α that minimizes the functional $G_\lambda(q^\alpha)$ given by

$$q^\alpha = \underset{q^\alpha}{\operatorname{argmin}} \, ||G_\lambda(q^\alpha)||_{\mathcal{D}}. \tag{2.5}$$

It now remains to find the appropriate regularization parameter and operator. There are several methods proposed in literature on how to do this. For example, the operator F can be found by using Fourier, Laplace, and other forms of integral transformations [6]. Also, the regularization parameter can be determined by using *L-curve* [6, 36]. Most commonly, in the case of Tikhonov regularization scheme, a smoothening function $g : \mathcal{Q} \to \mathcal{D}$ is constructed and added to the original operator. Hence, the regularization operator becomes

$$G_\lambda^\alpha = G_\lambda + \lambda g_\lambda. \tag{2.6}$$

Usually, g_λ is chosen as a Tikhonov functional. In practice, the second order Tikhonov functional is chosen for most problems [85]. Thence, (2.5) is transformed to finding q^α that satisfies

$$q^\alpha = \underset{q^\alpha}{\operatorname{argmin}} \, ||G_\lambda^\alpha(q^\alpha)||_{\mathcal{D}}. \tag{2.7}$$

It is worth noting that Tikhonov functional is weighted *Sobolev norm* associated with the *Sobolev space* W_2^p. Therefore, finding q^α involves looking for solutions that minimize the Sobolev norm of the functional $G_\lambda^\alpha(q^\alpha)$.

2.3 Modeling and Formulation of Local Volatility

In this section, we discuss the local volatility model in greater detail. We start off with the Black-Scholes model which gives an analytical formula to calculating European option prices. Subsequently, we will discuss some of the assumptions of the Black-Scholes model that do not hold in the real markets. So, we introduce the local volatility model as an alternative model to relax some of those assumptions to reflect market reality. Furthermore, several forms of local volatility equation that are well documented will be given. Also, the assumptions and

16

properties needed for the calibration of the local volatility model will be high-lighted. Consequently, we will give a more detailed analysis of the challenges facing the calibration of local volatility model.

Consider the model proposed by Fischer Black and Myron Scholes [14] for pricing European options. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{P})$ be a filtered probability space such that $\mathcal{F}_t$ is the natural filtration for W_t over the sample space Ω. Consider an asset S whose dynamics is driven by a geometric Brownian motion $(W_t)_{(t \geq 0)}$ under the objective probability measure $\mathbb{P}$. Then, the asset price S satisfies the stochastic differential equation (SDE) given by

$$dS_t = \mu S_t dt + \sigma S_t B_t, \tag{2.8}$$

where μ is the drift term of the asset price movement, and σ is the diffusion coefficient representing the volatility. The coefficients μ and σ are taken to be constant in the original Black-Scholes model. However, their results have been extended for the case where μ and σ are taken to be deterministic functions of time.

For the model to be useful in the market with non-arbitrage opportunities, there is need to transform the above SDE to follow a Brownian process driven by the risk-neutral probability measure $\mathbb{Q}$. This transformation is done using the *Girsanov theorem* which can be found in standard Stochastic Calculus textbooks (for example see [64, 81]). The new filtered probability space is given by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{(t \geq 0)}, \mathbb{Q})$ while the asset S follows the SDE

$$dS_t = r S_t dt + \sigma S_t W_t, \tag{2.9}$$

where r is the risk-free rate in the market. Given (2.9), the option price V written on an asset S over time t satisfies the Black-Scholes partial differential equation (BS-PDE)

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0, \tag{2.10}$$

with zero dividends. Solving the above PDE gives the European price of the option. Black and Scholes analytically solve (2.10) to give the famous Black Scholes formula given in the theorem below:

Theorem 2.2. *Given the assumptions and the dynamics imposed on S in* (2.9) *the European call option price in the time horizon $t \in [0, T]$ is given by*

$$C_{BS}(t, S_t; T, K, \sigma_{BS}, r) = S_t N(d_1) - K e^{-r(T-t)} N(d_2), \qquad (2.11)$$

with $d_1 = \frac{\ln(\frac{S_t}{K}) + \frac{\sigma_{BS}^2 T}{2}}{\sigma_{BS}\sqrt{T}}$ and $d_2 = d_1 - \sigma_{BS}\sqrt{T}$, where K is the strike price and T is the maturity of the option.

For a full proof of Black-Scholes model, readers should refer to [54]. As shown in (2.11), to price an option, the parameters S_0, K, T, r, σ have to be specified. However, in practice all the other parameters except σ can directly be observed from the market. Therefore, in order to specify a volatility σ to the model, two approaches are often used. First approach uses the historical information of the asset price movements to estimate the volatility. The second approach involves inversion of the Black-Scholes price to find the unique volatility, the so called *implied volatility* which we have denoted as σ_{BS}. The latter approach is more suitable for calibration purposes because it is forward looking compared to the former.

We also note that the implied volatility for different options written on an asset S_0 at time $t = 0$, varies for each option price $V(K, T)$ specified by strike K and maturity T (see for example [73, 77]). This creates a one-to-one relationship between the option's implied volatility and its price. However, this contradicts the assumption that the volatility for the asset S_0 is unique for all options written on it. This leads to the conclusion that the volatility must be dependent on the strike K and maturity T of the option. We call this new volatility, the local volatility denoted by $\sigma_{LV}(K, T)$. The dependence of volatility on the strike K is called the volatility *smile* or *skew* while the variation with respect to maturity T is regarded as the *term structure* of volatility. Without accounting for skewness in the structure of the volatility as in the case of Black-Scholes model, it would suggest that every option price, V_t, $t \geq 0$, written on an asset, S_t, follows a different price process, which is clearly not the case. Furthermore, it would suggest that when building binomial trees for option pricing, different trees would need to be constructed for the asset process each time.

As explained by Animoku et al. [5], the local volatility model among all other

models tries to model volatility skewness in a less computationally intensive way. In the next section, we turn our attention to describing different forms of local volatility models proposed in the literature.

2.3.1 Dupire Local Volatility Equation

In the original paper by Dupire [33], he showed that the local volatility surface of an asset can be uniquely specified given observable market prices for European options for all strikes and maturities. Ever since, several forms of the original local volatility equation have been derived in the literature. In this section, we set up the local volatility model and its various functional forms, see for example [5, 39]. First, we setup the model as follows: let $(\Omega, \mathbb{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{Q})$, be a given filtered probability space where $\mathbb{Q}$ is the risk neutral probability measure defined on the given probability space. Consider a local volatility model, in which the risky asset price, S_t, follows the dynamics

$$dS_t = \mu_t S_t dt + \sigma(t, S_t) S_t dW_t, \tag{2.12}$$

where $\mu_t = r_t - q_t$, and W_t is the Brownian motion under the risk-neutral probability measure $\mathbb{Q}$. Furthermore, $\mu : \mathbb{R}^+ \to \mathbb{R}$, and $\sigma : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ are continuous functions such that for $t \geq 0$, and $(x, y) \in \mathbb{R}^2$,

$$|x\sigma(t, x) - y\sigma(t, y)| \leq M |x - y|$$

holds for any $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$ and $\sigma(t, x) \geq m$, where m and M are positive constants. The natural filtration of $(W_t)_{t \geq 0}$ is denoted by $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$.

Here, we focus on local volatility equations derived under European option prices. Although we start off with a volatility of the form $\sigma(t, S_t)$, which depends on time and asset level, we pass to a more convenient way of characterizing the volatility, specifically $\sigma(T, K)$, depending on maturities and strike prices. Note that we have restricted the notations for the parameters of the option prices to include only maturity T and strike K for simplicity.

Next, we present the fundamental theorems in the literature on characterizing the local volatility function. Proofs are omitted and we refer to, for instance, [39].

Theorem 2.3 (Dupire Local Volatility Equation). *Let S be the dynamics of the asset defined in (2.12) and $C(T, K)$ be the European call option with strike K and maturity T. Then, we have*

$$\sigma^2(T, K) = \frac{\left(2\left[\frac{\partial C(T,K)}{\partial T} + K(r_T - q_T)\frac{\partial C(T,K)}{\partial K} + q_T C(T, K)\right]\right)}{\left(K^2 \frac{\partial^2 C(T,K)}{\partial K^2}\right)} \tag{2.13}$$

for $T > 0$ and $K > 0$.

For computational efficiency, we work with Dupire equation under the forward price rather than spot prices. Therefore, we present the local volatility under the forward price as well. Let $C(T, K) = C(F_T, T, K, \sigma_{LV}(T, K))$ and $F_t = F(t, T) = S_t e^{\int_t^T \mu_s ds}$ so that $F_T = S_T$. Writing the SDE in (2.12) under the dynamics of the forward price yields

$$\begin{aligned} dF_t &= e^{\int_t^T \mu_s ds} dS_t - \mu_t S_t e^{\int_t^T \mu_s ds} dt \\ &= e^{\int_t^T \mu_s ds}\left(\mu_t S_t dt + \sigma(t, S_t)S_t dW_t\right) - \mu_t F_t dt \\ &= \tilde{\sigma}(t, F_t)F_t dW_t, \end{aligned} \tag{2.14}$$

where $\tilde{\sigma}(t, x) = \sigma(t, xe^{-\int_t^T \mu_s ds})$.

Theorem 2.4 (Local Volatility in terms of Forward Price). *Let F_t be the dynamics of the forward price of the asset given in (2.14) for $t \geq 0$ and $C(T, K)$ be the European call option with strike K and maturity T expressed in terms of forward price. Then, the following equation holds:*

$$\sigma^2(T, K) = \frac{\frac{\partial C(T,K)}{\partial T}}{\frac{1}{2}K^2 \frac{\partial^2 C(T,K)}{\partial K^2}}. \tag{2.15}$$

Equations (2.13) and (2.15) are termed as the Dupire local variance under spot and forward price dynamics, respectively.

Finally, we present the local volatility equation in terms of implied volatilities. This is necessary since in some financial markets, implied volatilities are quoted rather than the option prices themselves.

Lemma 2.5. *Let the Black-Scholes total variance be*

$$w(S_0, T, K) = \sigma_{BS}^2(S_0, T, K)T$$

and $y = \log\left(\frac{K}{F}\right)$. *Then, the Black-Scholes price in terms of w and y is*

$$C_{BS}(F, w, y) = FN(d_1) - Fe^y N(d_2)$$

with $d_1 = \frac{-y}{\sqrt{w}} + \frac{\sqrt{w}}{2}$ and $d_2 = \frac{-y}{\sqrt{w}} + \frac{\sqrt{w}}{2} - \sqrt{w}$.

Theorem 2.6 (Local Volatility as a Function of Implied Volatility)**.** *If the call option written on an asset is described by the Black-Scholes formula in Lemma 2.5, then the local variance of the asset can be obtained in terms of w and y as*

$$\sigma^2(T, K) = \frac{\frac{\partial w}{\partial T}}{\left[1 - \left(\frac{y}{w}\right)\frac{\partial w}{\partial y} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{w} + \frac{y^2}{w^2}\right)(\frac{\partial w}{\partial y})^2 + \frac{\partial^2 w}{2\partial y^2}\right]}. \tag{2.16}$$

Immediate corollary of the theorem is given when $\frac{\partial w}{\partial y} = 0$:

$$\sigma^2(T, K) = \frac{\partial w}{\partial T}, \text{ or equivalently, } w(T) = \int_0^T \sigma^2 ds$$

for maturity T.

For the proofs of the above theorems, interested readers should see [39] for example, and the references therein.

2.3.2   Properties of Local Volatility Model

In order to calibrate the various forms of local volatility model presented in the previous section, we need to specify some properties that the surfaces should exhibit. These properties will be important when calibrating under the Bayesian framework as they will be incorporated in to the prior distribution. These properties include:

(a) *Positivity*: In order to implement Dupire's equation, we adopt $\sigma_{LV} > 0$.

(b) *Asymptotic*: The local volatility values at time $t_0 = 0$ should be fitted to the implied volatilities σ_{imp} for quoted European options. Thus, the implied volatilities at $t_0 = 0$ determine the local volatility surface completely in $\mathbb{R}^3$; see [10].

(c) *Smoothness*: We also ensure smoothness of the local volatility surfaces. This regularizes the surfaces so that spikes or troughs are not observed. Since it is not justified to expect a volatility surface to predict abrupt movements in future volatility [55, 45].

2.3.3   Challenges of Local Volatility Model

With the various forms of local volatility equation shown in Section 2.3.1, local volatility surfaces can be constructed from the option prices and implied volatilities observed in the market. A crucial assumption is that the option price $C(T, K)$ is considered to be of class $\mathcal{C}^{1,2}$. That is, the map $(T, K) \mapsto C(T, K)$ is continuously differentiable with respect to the first argument T and twice continuously differentiable with respect to the second variable K. However, even if this assumption holds, there is still a problem with option price function being unknown analytically which causes problems when taking partial derivatives. Therefore, the partial derivatives of the call function have to be estimated numerically. Due to the imperfect nature of numerical methods, the algorithm used in estimating local volatility function may be unstable. That is, small changes in the input data may result in large error in the function values. Observe that in the denominator of (2.13), small errors in the partial derivative can be magnified by the square of the strike price. This can lead to negative values of local volatility, which is unacceptable.

On the other hand, the continuity assumption of option prices is unrealistic. In practice, limited number of option values are known for finite number of maturities and strike prices which makes the local volatility equation ill-posed. To solve this problem, one can smoothen the option price data using Tikhonov regularization or by minimizing the function's entropy [24, 48]. Another viable method is to use smoothing cubic spline interpolation to obtain arbitrage-free option prices [37]. These methods should be able to guarantee the convexity of the option prices in the strike direction which adds extra complexity to the model. Also, the call option function must be monotonically decreasing in strike and increasing in maturity to avoid calender arbitrage. This way, arbitrage-free

prices can be ensured.

2.4 Error Minimization Method of Solving Dupire Local Volatility

In this method, let $V(S, t; T, K, \sigma)$ be the option price at time t and $\sigma(t, S)$ be the choice of volatility function, then the option prices, V, follow the stochastic Black-Scholes PDE in (2.10). Thus, given the initial and boundary conditions of (2.10), the option prices, $V(S_0, 0; T, K, \sigma)$, can be solved uniquely. Since we are dealing with standard European options, the appropriate initial and boundary conditions are;

$$
\begin{cases}
V(S, T; T, T, \sigma) = \max(S_T - K, 0) & \text{for } S \geq 0, \\
V(0, t; T, K, \sigma) = 0 & \text{for } 0 \leq t \leq T, \\
\frac{\partial V}{\partial S}(S, t; T, K, \sigma) = e^{-q(T-t)} \ as \ S \to \infty & \text{for } 0 \leq t \leq T.
\end{cases}
\tag{2.17}
$$

In the general setup above, market calibration involves finding a local volatility function σ that solves the PDE in (2.10) such that the obtained option prices fall in between the corresponding bid and ask spread. That is,

$$
V_{ij}^b \leq V(S_0, 0; T, K, \sigma) \leq V_{ij}^a,
$$

for $i = 1, 2, 3, \ldots, N$ denoting the sets of maturities, T_i's and $j = 1, 2, 3, ..., M$ denoting the sets of strike prices, K_j's, for each maturity. Satisfying these inequality constraints, a function $G(\sigma)$ is to be minimized with respect to σ and possibly making it approach zero:

$$
G(\sigma) = \min_{\sigma \in \mathcal{Q}} \sum_{i=1}^{N} \sum_{j=1}^{M_i} [V(S_0, 0; T_i, K_{ij}, \sigma) - \tilde{V}_{ij}]^2,
\tag{2.18}
$$

where $\tilde{V}_{ij} = \frac{1}{2}(V_{ij}^a + V_{ij}^b)$ is the average of the bid and ask prices. To this extent, minimizing the function G over a general space of admissible functions is ill-posed, essentially because we have a finite and discrete number of observable option prices. Hence, the function σ can not be uniquely determined with guaranteed continuous dependence on market option prices. As a consequence, a small perturbation in price data can lead to a large change in the minimizing function.

To overcome this challenge, a Tikhonov regularization scheme is specified in the literature. For more details on the application of Tikhonov regularization in solving Dupire local volatility equation, interested readers should see [1, 34, 48, 63] to mention a few.

2.5 Local Volatility Calibration: Literature review

In the literature several authors have proposed different error minimization techniques in solving the Dupire local volatility equation. Most commonly used methods involve the use of Tikhonov regularization in solving the non-linear problem. However, according to our best knowledge little work has been done in the use of Bayesian technique to characterize the distribution of the calibrated local volatility parameter. Herewith, we give comprehensive literature review on the results obtained so far on local volatlity modeling and the numerical methods used in its implementation.

First, we discuss some of the literature involving the use of parametric methods in constructing the local volatility surface. Cerrato [17] uses the Dumas et al parametrization [32] of the implied volatility in calibrating the local volatility parameter. Here, the author uses the original form of Dupire local volatility equation in (2.13). Although the computational procedure used in [17] is quite efficient and fast, the numerical accuracy is lacking. The local volatility surface contains spikes and is generally flat which suggests the resultant surface is lacking robustness. Moreover, the model prices at some discretized points are negative, which poses serious confidence on the accuracy of the numerical procedure.

More recently, Animoku et al. [5] similarly used parametric method in constructing local volatility surfaces. However, their implementation is different from [17] in two ways. First, the choice of implied volatility parametrization involves the use of moneyness instead of strike price, which provides alternative interpretation for the volatility surfaces. Secondly, the authors use the local volatility function in (2.13) for their numerical implementation, in addition to (2.16) used in [17].

In the remainder of this section, we discuss the non-parametric methods of obtaining the local volatility surfaces that have been implemented in the literature. First, we discuss the method used by Lagnado & Osher [63]. In their methodology, the gradient of the local volatility functional (in the form of (2.7)) is minimized in $\mathcal{L}_2$ norm over an appropriate space of smooth or cost functions subject to constraints that ensure the solution of the pricing PDE are within the bid-ask spread of the observed prices. The authors use the error minimization approach explained in the previous section, however to ensure an optimal solution that do not violate the Hadamard conditions, Tikhonov regularization is used. Therefore, the authors minimized the following functional

$$G_1(\sigma) = G(\sigma) + \lambda \left\| \left\| \nabla \sigma \right\| \right\|_2^2, \tag{2.19}$$

where the regularizing parameter λ is a constant chosen by trial and error in order to optimize the convergence rate of the minimization procedure. The authors obtained an optimal solution that is numerically robust and close to minimizing $G(\sigma)$. However, their optimal solution to (2.19) is not proven analytically but shown to give good calibrated result for the market data taken from S&P 500.

The authors used gradient descent scheme to find the optimal σ, where they solve the Black-Scholes PDE (2.10) via finite difference method at each iteration. There are some drawbacks to the authors' computational procedure and the implementation results. On the computational aspect, their numerical procedure involves solving the PDE in (2.10) at every iteration which is computationally expensive. On the numerical result side, the resultant local volatility surface $\sigma(t, S)$ is only estimated for some specified discretized points which makes it difficult to price options out of this range.

Chiarella et al. [19] proposed a more computationally efficient method to reduce the complexity of the algorithm used in [63] by estimating the parameter σ using the Black-Scholes formula in (2.11). The authors' method is computationally fast and involves few iterations at each step. However, the surfaces produced using their numerical scheme still lacks the details necessary for pricing exotic options that depend on out-of-the money volatility values.

Jackson et al. [55] presented a more direct approach of minimizing the error

functional in (2.19) which avoids the computation of variational derivatives. The authors first represent the local volatility surface by discretizing the parameter space into a set of nodes, which are later interpolated using natural cubic splines. In their calibration, greater weights are chosen for at-the money options compared to out-of-the money options so that the contributions of each option to the volatility surface is reflected. The authors use a quasi-Newton algorithm for the optimization procedure and a piecewise quadratic finite element method for solving the Black-Scholes PDE (2.10) at each iteration.

This procedure reduces the computational time drastically, however, it has a number of drawbacks. First, the regularization parameter λ is still chosen arbitrarily by trial and error. Even though λ is chosen to improve the convergence of the algorithm, it lacks financial interpretation. Secondly, the volatility parameter is only calibrated for relatively few node points (15 nodes) for 10 option prices. However, in practical applications, we usually expect to calibrate for more than 100 prices, which can not be easily done with the authors proposed numerical scheme. Thirdly, the method presented is vulnerable to over-regularization as they seek to obtain a unique solution. That is why we observe that the pricing errors in their article are relatively high; for option values between 4 to 7 basis points.

Now, we turn to some of the theoretical results that have been obtained concerning stability and convergence analysis in minimizing the error functional presented in (2.19). For example, Crepey [23] has used Tikhonov regularization to prove the stability and convergence of the local volatility estimates. In his approach, he specifies a deterministic prior function σ_0 together with a regularization term and seeks to minimize the difference between this and the calibrated σ. It is worth noting that the prior guess in his method is not in a probabilistic sense as demonstrated by the Bayesian method. Therefore, the author's proposed error functional is given by

$$G_2(\sigma) = G(\sigma) + \lambda \left(\|\sigma - \sigma_0\|_2^2 + \||\bigtriangledown \sigma|\|_2^2 \right). \tag{2.20}$$

Furthermore, authors such as Egger & Engl [34] have also minimized the functional (2.20) by specifying a different prior guess for σ.

Egger et al. [35] uses a much different approach to the methods we have discussed so far by decoupling the volatility into the smile and term structure, so that we can express the local volatility as

$$\sigma(t, S) = \sigma_1(t)\sigma_2(S). \tag{2.21}$$

Although, the authors approach is quite unique, it is numerically challenging to use as a general framework in finding the local volatility function. This is because the method is not practical for market data whose volatility smile is inconsistent over time.

Coleman et al. [20] used a bicubic spline to approximate the local volatility surface by matching the spline knots to coordinates (K, T) of the observable market prices. In order to reduce the space of solutions, they placed a bound on the volatility values at each knot. The computational cost of the authors method is considerable given that they calibrated 70 spline knots to 70 market prices.

Now, we discuss some of the recent work on the use of Bayesian analysis in constructing local volatility surfaces. In a more recent study, Gupta & Reisinger [45] reformats the inverse problem of calibrating local volatility parameter into a Bayesian framework where a distribution of surfaces for the calibrated parameter were obtained. Consequently, the convergence of the procedure was also monitored for the MCMC chains. In the authors' methodology, the errors from the statistical model are modeled as additive and normally distributed. The consistency of the volatility surfaces over time were also examined. The drawback in their method is of statistical nature. The additive errors from the authors' method, when graphed shows some dependency on the market option prices which suggest a multiplicative error statistical model is more suitable for the calibration. Not withstanding, the results from their numerical procedure shows good calibrated volatility surfaces that are robust and consistent.

As a follow up work on [45], Animoku et al. [5] explored both additive and multiplicative statistical models in the volatility parameter calibration using Bayesian framework. The errors were modeled as both normal and student-t distributions for each statistical model. The numerical result from the authors'

implementation, shows a smooth and robust volatility surface with expected skewnees at maturities close to $t = 0$.

CHAPTER 3

FREQUENTIST APPROACH TO PARAMETER ESTIMATION

As discussed in the previous chapters, most inverse problem involves parameter estimation in a robust manner given some noisy observations. From *frequentist* perspective, such parameters are assumed to be unknown but not random unlike the Bayesian setting treated in Chapter 4. This involves estimating a calibrated parameter in a deterministic way. Thus, we consider the statistical model

$$\Upsilon = f(q_0) + \epsilon, \tag{3.1}$$

where $\Upsilon = [\Upsilon_1, \ldots, \Upsilon_n]^T$ is a random vector representing the observations or measurements from experiment whose realization is $\upsilon = [\upsilon_1, \ldots, \upsilon_n]^T$. The random vector $\epsilon = [\epsilon_1, \ldots, \epsilon_n]^T$ represents the measurement errors. The errors are additive, that is, the errors represent the addition of the model and measurement errors.

The aim of the frequentist approach is to estimate the parameter q so that the model response fits the data in some optimal sense. This can be done by constructing an estimator $\hat{q}$ such as ordinary least squares (OLS) and maximum likelihood estimators that statistically estimate q_0 in a reasonable sense. For example, the OLS estimator is given by

$$\hat{q} = \operatorname*{argmin}_{q \in \mathcal{Q}} \sum_{i=1}^{n} \left[\Upsilon_i - f_i(q) \right]^2.$$

Furthermore, since the estimator $\hat{q}$ is a random variable or random vector it has statistical properties such as the mean, covariance and distribution termed as

the *sampling distribution.* Thus, using the expectation and sample covariance of the estimator, one can quantify the variability of the errors in order to validate the assumptions.

The sampling dstribution however does not provide a distribution for the model parameters since q_0 is not random in frequentist inference. However, in certain applications the sampling distribution can coincide with the posterior distribution from Bayesian techniques.

The rest of the chapter is setup as follows: In Section 3.1, we focus on deriving the sampling distribution under linear regression model. Consequently in Section 3.2, the sampling distribution under the non-linear case is studied. Finally, the numerical implementation of frequentist approach in constructing local volatility surfaces is discussed in Section 3.3.

3.1   Linear Regression

We highlight some fundamental results regarding linear regression here to motivate similar theoretical results for the nonlinear least squares case. For more details, we refer the reader to [44, 75]. We summarize the main results here.

Consider the statistical model

$$\Upsilon = X q_0 + \epsilon, \tag{3.2}$$

where $\Upsilon = [\Upsilon_1, \ldots, \Upsilon_n]^T$, $\epsilon = [\epsilon_1, \ldots, \epsilon_n]^T$, and $n \times p$ design matrix X is deterministic and known. Let q_0 denote the vector of true but unknown parameters and let $\upsilon = [\upsilon_1, \ldots, \upsilon_n]^T$ denote observations of an experiment of which the errors are $\epsilon = [\epsilon_1, \ldots, \epsilon_n]^T$. We assume $n > p$.

First, we specify the assumptions around the distribution of the errors for the statistical model in (3.2).

Assumption 3.1 ([75], Assumption 7.2). *Assume that the errors are unbiased and i.i.d. with fixed but unknown variance σ_0^2; then, for $i = 1, \ldots, n$,*

- $\mathbb{E}[\epsilon_i] = 0$;

- $\mathbb{V}ar[\epsilon_i] = \sigma_0^2,\ \mathbb{C}ov[\epsilon_i, \epsilon_j] = 0\ \ for\ \ i \neq j.$

The assumptions above are in accordance with the frequentist approach. Next, we construct the unbiased estimators $\hat{q}$ and $\hat{\sigma}^2$ for unbiased parameters q_0 and σ_0^2. To construct an unbiased estimator $\hat{q}$ for q_0, we seek to find q which minimizes the OLS functional

$$\mathcal{J}(q) = (\Upsilon - Xq)^T(\Upsilon - Xq).$$

Taking the gradient and equating to zero gives

$$\bigtriangledown_q \mathcal{J}(q) = 2\left[\bigtriangledown_q(\Upsilon - Xq)^T\right][(\Upsilon - Xq)] = 0, \tag{3.3}$$

where $\bigtriangledown_q(\Upsilon - Xq)^T = -\bigtriangledown_q q^T X^T = -X^T$. This leads to the least squares estimator

$$\hat{q}_{OLS} = (X^TX)^{-1}X^T\Upsilon, \tag{3.4}$$

and its realization

$$q_{OLS} = (X^TX)^{-1}X^Tv. \tag{3.5}$$

Proposition 3.2 (Properties of the Estimators, [75], Results 7.4). *The parameter estimator $\hat{q}$ has the mean and covariance matrix as shown below:*

1. *$\mathbb{E}[\hat{q}] = q_0$, which can easily be shown by*

$$\mathbb{E}[\hat{q}] = \mathbb{E}\left[(X^TX)^{-1}X^T\Upsilon\right] = (X^TX)^{-1}X^T\mathbb{E}\left[\Upsilon\right] = q_0.$$

   Thus, q provides an unbiased estimate for the true parameter q_0.

2. *$\mathbb{V}ar[\hat{q}] = \sigma_0^2(X^TX)^{-1}$. For the covariance relation, let $A = (X^TX)^{-1}X^T$, then,*

$$\begin{aligned}
\mathbb{V}ar[\hat{q}] &= \mathbb{E}\left[(\hat{q} - q_0)(\hat{q} - q_0)^T\right] \\
&= \mathbb{E}\left[(q_0 + A\epsilon - q_0)(q_0 + A\epsilon - q_0)^T\right]\ \ since\ \ \hat{q} = A\Upsilon = A(Xq_0 + \epsilon) \\
&= \mathbb{E}\left[(A\epsilon)(A\epsilon)^T\right] \\
&= A\mathbb{E}\left[(\epsilon\epsilon^T)\right]A^T \\
&= \sigma_0^2(X^TX)^{-1},
\end{aligned}$$

   with σ_0^2 assumed to be fixed as noted earlier.

3. For the variance estimator, we have

$$\hat{\sigma}^2 = \frac{1}{n-p} \hat{R}^T \hat{R}, \tag{3.6}$$

where $\hat{R} = \Upsilon - X\hat{q}$.

We omit the proof for the covariance estimator here, interested reader should check [75, 44].

Given the results above, we can determine the sampling distribution for the parameter q. With this distribution, one can also construct the confidence intervals and compute some meaningful statistics for this distribution.

Assumption 3.3. *The sampling distribution $\hat{q}$ can be specified for problems with errors assumed to have properties shown in Assumption 3.1.*

Corollary 3.4 (Sampling Distribution for $\hat{q}$, [75], Property 7.8). *For $\epsilon \overset{iid}{\sim} \mathcal{N}(0, \sigma_0^2)$, the sampling distribution for $\hat{q}$ is given as,*

$$\hat{q} \sim \mathcal{N}\left(q_0, \sigma_0^2 (X^T X)^{-1}\right).$$

Furthermore, if we let the δ_k, $k = 1, \cdots, p$ denote the kth diagonal element of $(X^T X)^{-1}$ and q_{0_k} denote the kth element of the true parameter vector q_0, then

$$\hat{q}_k \sim \mathcal{N}\left(q_{0_k}, \sigma_0^2 \delta_k\right).$$

Algorithm 3.1 Algorithm for the Frequentist approach to parameter estimation

Model:

$$\Upsilon = Xq_0 + \varepsilon, \quad q \in \mathbb{R}^p,$$

$$\upsilon = Xq_0 + \epsilon \quad \text{(realization)}.$$

Assumptions: $\mathbb{E}(\varepsilon_i) = 0$, ε_i is i.i.d. with $\mathbb{V}(\varepsilon_i) = \sigma_0^2$.

Least Squares Estimator and Estimate:

$$\hat{q} = (X^T X)^{-1} X^T \Upsilon, \quad = \mathbb{E}[\hat{q}] = q_0, \quad \mathbb{V}(\hat{q}) = \sigma_0^2 (X^T X)^{-1},$$

$$q = (X^T X)^{-1} X^T \upsilon.$$

Error Variance Estimator and Estimate: $\hat{R} = \Upsilon - X\hat{q}, \quad R = \upsilon - Xq$

$$\hat{\sigma^2} = \frac{1}{n-p} \hat{R}^T \hat{R}, \quad \hat{\sigma^2} = \frac{1}{n-p} R^T R.$$

Covariance Matrix Estimator and Estimate:

$$\mathbb{V}(\hat{q}) = \hat{\sigma^2}(X^T X)^{-1}, \quad \mathbb{V} = \sigma^2 (X^T X)^{-1}.$$

Sampling Distribution: Requires $\varepsilon \sim \mathbb{N}(0, \sigma_0^2)$ or sufficiently large n

- $\hat{q} \sim \mathcal{N}\left(q_0, \sigma_0^2 (X^T X)^{-1}\right)$;

- $(1-\alpha) \times 100\%$ Confidence Intervals: $\delta_k = [(X^T X)^{-1}]_{kk}$

$$q_k - t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k}, q_k + t_{n-p, 1-\alpha/2} \sigma \sqrt{\delta_k}.$$

3.2 Non-linear Regression

Non-linear problem takes the following statistical form:

$$\Upsilon = f(q_0) + \varepsilon, \tag{3.7}$$

where $f(q_0)$ is the model response and q_0 is the true but unknown parameter to be quantified. As before, we assume that there are more observations than parameters, that is, $n > p$. Let Q be the support for the parameter $\hat{q}$ from the admissible parameter space $\mathbb{Q}$.

Similar to the construction of the OLS estimate in Section 3.1, we also minimize the functional

$$\mathcal{J}(q) = \sum_{i=1}^{n} [v_i - f_i(q)]^2 \tag{3.8}$$

subject to $q \in \mathbb{Q}$.

Unlike the results obtained in Section 3.1, it is quite difficult to obtain an analytical expression for the minimizer in (3.8). Instead, we seek a numerical estimate that minimizes the least squares functional. The algorithm for the nonlinear case is similar to that of Algorithm 3.1 with some adjustments. We replace the statistical model in the linear case with (3.7) and the variance of the estimator $\hat{q}$ with

$$\mathbb{V}(\hat{q}) = \sigma_0^2 \left[\mathcal{X}(q_0)^T \mathcal{X}(q_0) \right]^{-1} \approx \hat{\sigma}^2 \left[\mathcal{X}(q)^T \mathcal{X}(q) \right]^{-1}.$$

$\mathcal{X}(q)$ denotes the $n \times p$ sensitivity matrix whose entries are

$$\mathcal{X}_{ik}(q) = \frac{\partial f_i(q)}{\partial q_k}.$$

Note that the variance is constructed here by linear approximation around q_0. There are three methods that can be used to construct the sensitivity matrix, namely: finite difference approximations, solution of sensitivity equations, and automatic differentiation. For further details on how to use these methods to construct sensitivity matrices, see [75]. We should note that we have used the finite difference approximations in constructing the sensitivity matrix used in approximating the variance-covariance matrix for the local volatility parameter. For more details on non-linear regression, interested readers should see [7, 8, 60].

3.3 Implementation of Frequentist approach in Constructing Local Volatility Surfaces

In this section, we focus on estimating the local volatility parameter via the frequentist approach. Since we are dealing with a non-linear model, we seek non-linear estimates from the sampling distribution of the parameter σ_{LV}. At time $t = 0$, we expect the sampled estimates of the parameter to converge to the implied volatility. Furthermore, the variance-covariance (sensitivity) matrix of the parameter is numerically computed via the method explained in Section 3.2. As a result, the sampling distribution of the parameter is given by

$$\sigma_{LV} \sim \mathcal{N}\left(\sigma_{imp}, C\right), \tag{3.9}$$

where σ_{imp} is the implied volatility and C is the sensitivity matrix of the parameter. The estimates of the parameter, σ_{LV}, can be sampled from its sampling distribution by

$$\hat{\sigma}_{LV} = \sigma_{imp} + Rz,$$

where R is the *Cholesky* decomposition of the covariance matrix C, and z is standard normal random variable. Using the set of data specified in Appendix B together with Algorithm 3.1, we compute the local volatility surface shown in Figure 3.1 from the estimates computed in (3.9).

Here, we take a sample of 10000 estimates of $\hat{\sigma}_{LV}$ and compute its expectation. The expected value is then used to construct the volatility surface. The result shows that the sums of squared errors for the local volatility surface under the frequentist approach is 1701.7.
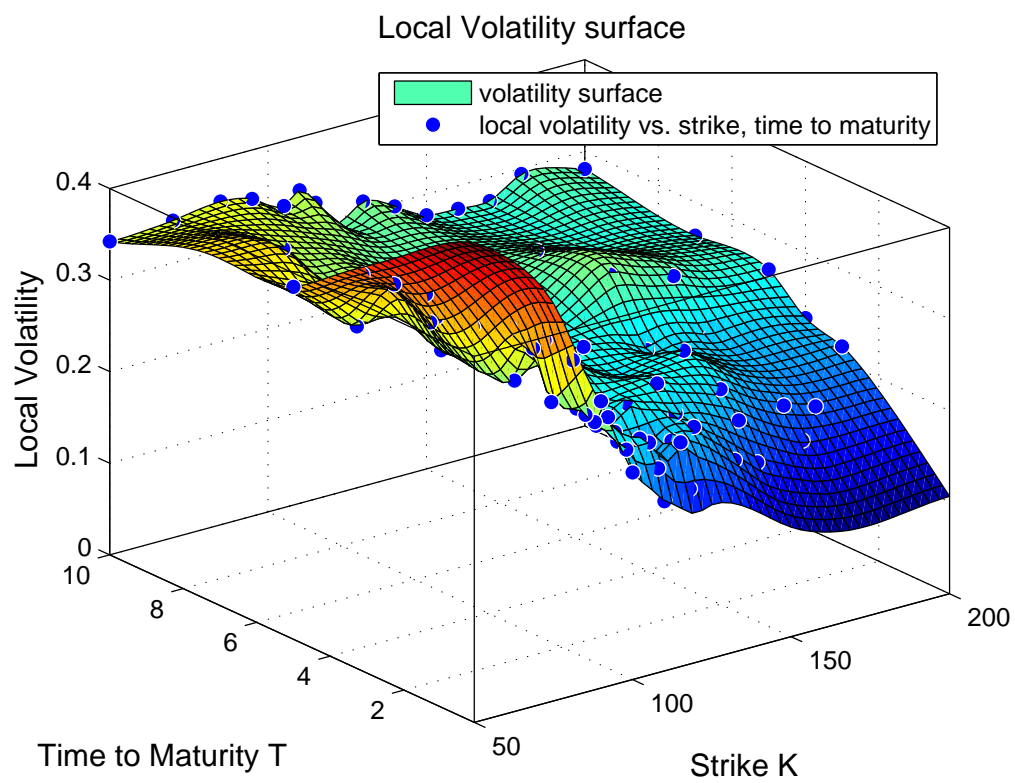
Figure 3.1: Local volatility surface via frequentist approach

CHAPTER 4

BAYESIAN MODELING OF LOCAL VOLATILITY

In this chapter, we introduce the theory as well as numerical implementation of Bayesian inference in parameter estimation. Bayesian method involves finding a probability distribution for an uncertain parameter in a statistical model. The so called *posterior distribution* of the parameter. An important aspect of constructing the Bayesian posterior distribution is incorporating a prior set of beliefs about the parameter to be estimated. These prior information are incorporated mathematically into the prior density, $\pi_0(Q)$, where Q is the set of parameter we wish to estimate.

Most advocates for Bayesian learning argue that omitting prior information in some significant amount of statistical problems often leads to inaccurate conclusions. However, the critics of this method have pointed out the importance of using accurate prior information to justify the use of Bayesian analysis. To that extent, it is unclear if the general adoption of Bayesian reasoning is justified. Although, from the Bayesian viewpoint, it is far more likely to recognize when a prior information is reliable and important in estimating a parameter, thus, it makes a totally sensible action to utilize such information rather than not. We will discuss the full implications of using Bayesian framework in Section 4.1.2.

Furthermore, there is need to construct the *likelihood function*, $L(v|q)$ under which the sampled data v are more likely to be observed given the parameter. The posterior density can then be constructed by combining the prior distribution and the likelihood function. This is demonstrated by the *Bayes' rule.*

The posterior density of the parameter reflects the best probability distribution of the parameter values based on the sampled data. In most applications, sampling parameter estimates from the posterior density is often difficult using the standard quadrature techniques. This is mostly due to the large number of parameters to be estimated in most inverse problems. A more useful technique is to sample these estimates as a Markov chain via a Metropolis Hastings algorithm. Due to the nature of the financial problem we seek to solve, we have employed this method judiciously in estimating the local volatility parameter via Bayesian framework.

The Bayesian method allows us to recast the inverse problem into a Bayesian framework. Thus, we compute estimates of the volatility by sampling from the posterior distribution. These estimates can then be used to obtain the entire volatility function using various interpolation techniques.

On one hand, the prior distribution involves a set of prior beliefs about the parameter to be estimated; and these prior knowledge are incorporated into the posterior distribution of the parameter. On the other hand, the likelihood function is constructed from the error distribution of the statistical model used in calibrating the model to the market prices.

The error distribution is chosen suitably to reflect the properties of the financial model used. In this thesis, we focus on two types of error distributions namely: *normal and student-t distributions*. We have assumed the model errors are distributed as normal or student-t for two reasons. First, these error distributions are very convenient to use in financial applications as many market parameters like logarithm of asset returns are normally distributed [18]. Second, given that the volatility of an asset is dependent on the asset return, it is natural to use normal error distributions in its calibration. In the case of a small sample size, student-t distribution is statistically preferable to normal distribution, thereby making the student-t distribution an appropriate choice for such case.

In this regard, we have tested the Bayesian method in estimating the local volatility parameter using four different statistical models involving both additive and multiplicative error models. In most engineering and financial applica-

tions, parameter estimation problems are often modeled as additive error given in the following statistical form

$$\Upsilon_i = f_i(Q) + \epsilon_i, \quad i = 1, \ldots, n, \tag{4.1}$$

where Υ_i, ϵ_i, and Q are random variables representing measurements, measurement errors, and parameters respectively; $f_i(Q)$ denotes the parameter-dependent model response. In addition, the errors are modeled as additive and mutually independent from Q. In most applications, the errors are usually modeled as normal and additive, see [45] for instance. We have shown in this work that such general statistical representation is often unreliable. This is collaborated by the fact that the plot of the errors shows a pattern indicating that the errors depend on the market data. Hence, we have extended the work in [45] to the multiplicative case where the errors depend on the magnitude of the sampled data. We expand on this in the later sections.

The remainder of this chapter is set up as follows. In Section 4.1, we give the theoretical set up for the Bayesian method in the context of parameter estimation. Here, we discuss the Bayes' rule and the various forms of Bayes' estimators. We conclude this section by identifying the advantages as well as the challenges of Bayesian inference in parameter estimation. In Section 4.2, we discuss the consistency of Bayes' estimators. Markov chains and convergence analysis for the Markov chain Monte Carlo method is explained in further details in Section 4.3. We formally present the four statistical models studied in Section 4.4. Furthermore, the mathematical constructions of the posterior densities for each statistical model are presented. Finally, in Section 4.5, we give some numerical results and recommendation based on the parameter estimates sampled from the posterior densities in Section 4.4.

4.1 Bayesian Framework

We recall that in constructing the posterior distribution of the parameter to be estimated, we need both the prior density $\pi_0(q)$ and the likelihood function $L(v|q)$. The prior density contains information that is previously known about

the parameter to be estimated. This information could be from previous experiments or observations. However, it is advised that caution be taken in choosing which prior knowledge to be incorporated. If the prior information is of questionable accuracy, it is advised that a non-informative prior, $\pi_0(q) = \mathbb{1}_{(0,\infty)}(q)$, be taken for positive parameters.

Furthermore, the likelihood function, $\pi(\upsilon|q) = L(q|\upsilon)$, contains the information that informs the posterior distribution given the sample observations. It also quantifies the conditional probability of obtaining the observed samples Υ given the values of the parameter Q. More mathematically, if $\pi(q,\upsilon)$ denotes the joint density of Υ and Q, then the likelihood is given by

$$\pi(\upsilon|q) = \frac{\pi(q,\upsilon)}{\pi_0(q)}.$$

For some observations $\upsilon = \upsilon_{obs}$, we can write the conditional probability above as

$$\pi(\upsilon_{obs}|q) = \frac{\pi(q,\upsilon_{obs})}{\pi_0(q)}.$$

Furthermore, the conditional probability of q given υ, which is termed as the posterior density can be written as

$$\pi(q|\upsilon_{obs}) = \frac{\pi(q,\upsilon_{obs})}{\pi(\upsilon_{obs})},$$

where we have assumed that

$$\pi(\upsilon_{obs}) = \int_{\mathbb{R}^p} \pi(q,\upsilon_{obs})dq = \int_{\mathbb{R}^p} \pi(\upsilon_{obs}|q)\pi_0(q)dq \neq 0.$$

Therefore, in the context of inverse problem, we seek to quantify the posterior density $\pi(q|\upsilon_{obs})$ given the set of observed data υ_{obs}.

The formal formulation of the inverse problem above depends on the *Bayes' theorem of inverse problems* which is explained in more details in [59]. We summarize the discussion in [59] below.

Theorem 4.1 (Bayes' Theorem of Inverse Problems, [75], Result 8.1)**.** *Suppose we have p random parameter variable Q with a known prior density $\pi_0(q)$. Suppose further that we have a set of observations υ_{obs} from the random observation Υ. Then, the posterior density of Q given the observed values υ_{obs} is*

$$\pi(q|\upsilon_{obs}) = \frac{\pi(\upsilon_{obs}|q)\pi_0(q)}{\pi(\upsilon_{obs})} = \frac{\pi(\upsilon_{obs}|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon_{obs}|q)\pi_0(q)dq}. \tag{4.2}$$

In general, construction of the likelihood function depends on the distribution of the errors. For instance, following the statistical model in (4.1), if we assume that the errors are i.i.d. and $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, where σ^2 is fixed, then the likelihood function can be specified as:

$$\pi(v|q) = L(q, \sigma^2|v) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-SS_q/2\sigma^2}, \tag{4.3}$$

where $SS_q = \sum_{i=1}^{n} [v_i - f_i(q)]^2$ is the sum of squared errors. It is worth noting that the construction of the likelihood here uses the probability density function of normally distributed random variable coupled with the i.i.d. property of the errors.

4.1.1 Bayes' Estimators

We have established that the Bayesian inference can be very useful in estimating an unknown parameter. These estimates are obtained via the probability distribution of the parameter. We seek the statistical properties of this distribution. For example, one can compute the point estimates of the parameter values such as the mean, median and mode. These Bayes' estimates can be constructed by attaching different loss functions to the posterior density.

Next, we define what a loss function is and various forms of Bayes' estimators. A function $H : \mathbb{R}^p \to \mathbb{R}$ is a *loss function* if and only if

$$H(q, q^{'}) = \begin{cases} 0, & \text{if } q^{'} = q, \\ > 0, & \text{if } q^{'} \neq q, \end{cases}$$

where $q, q^{'} \in \mathbb{R}^p$.

We expect the loss function to penalize estimates that are further from the true value compared to the ones closer to it. Thus, the larger the norm of the estimated value is from the true value the larger the loss function. So, we assume H is an increasing function of $\left\| q - q^{'} \right\|$ for some norm $\|.\|$ in $\mathbb{R}^p$. We now apply the loss function to the posterior distribution obtained in (4.2) to estimate values from the posterior density.

41

Definition 4.1. Given the loss function H, and the sample observations υ, the *Bayes' estimator* $q_H(\upsilon)$ constitutes the value of q that minimizes the expected loss with respect to the posterior. This is expressed mathematically as

$$q_H(\upsilon) = \underset{q'}{\mathrm{argmin}} \left\{ \int H(q, q') \pi(q|\upsilon) dq \right\}.$$

Remark 4.1. 1. The minimizer $q_H(\upsilon)$ is not necessarily unique

2. $H_1(q, q') = \left\| q - q' \right\|_2^2$ gives the *mean* of the Bayes' estimator $\bar{q} = q_{H_1}(\upsilon) = \mathbb{E}[q|\upsilon]$, with respect to the posterior density $\pi(q|\upsilon)$.

3. The uniform loss function $H_2(q, q') = \mathbb{1} - \mathbb{1}_{q=q'}$, the Bayes' estimator $q_{MAP} = q_{H_2}(\upsilon) = \max\{\pi(q|\upsilon)\}$; is called the *maximum a posteriori* (MAP) estimator, which is the mode of the distribution.

4. If $\pi_0(q)$ is uniform on $\mathbb{R}$, then q_{MAP} is equivalent to the *maximum likelihood estimate* q_{MLE} given by

$$q_{MLE} = \underset{q}{\mathrm{argmax}} \, \pi(\upsilon|q) = \underset{q}{\mathrm{argmax}} \, \pi(\upsilon|q)\pi_0(q) = q_{MAP}.$$

Once the parameter estimates have been sampled from a posterior distribution, one can perform some useful statistical analysis to further understand the calibrated parameter:

1. A *confidence interval* can be constructed around the *maximum a posteriori estimates* to understand the local behavior of the posterior using a Guassian or t-distribution. Suppose, we obtain q_0 as an estimate with standard deviation s, then $[q_0 - z_{\alpha/2}s, q_0 + z_{\alpha/2}s]$ and $[q_0 - t_{\alpha/2}s, q_0 + t_{\alpha/2}s]$ are the $(100 - \alpha)\%$ confidence interval assuming normal and student-t distributions, respectively.

2. *Marginal distributions* of each component of the parameter q_i, $i = 1, \ldots, p$, can be constructed from the joint densities. This can be easily done when using *Markov Chain Monte Carlo* method in sampling the estimates as the chain for each component of the parameter can be used to visualize the marginal distributions. Constructing the marginal distribution for each component is important in analyzing the contribution and sensitivity of each component to the joint distribution.

3. *Inferences* based on the behavior of another quantity of interest, that is a function of q, can be measured.

4.1.2 Advantages and Disadvantages of Bayesian Method

Here, we summarize the advantages and challenges of using Bayesian approach in parameter estimation in comparison to other regularization methods (Tikhonov regularization) discussed in Chapter 2.

First, the Bayesian method offers a natural way of choosing the regularization parameter λ as will be seen in Section 4.4. This is a crucial point because in Tikhonov regularization, λ is chosen ad-hoc through trial and error. In addition, the Bayesian framework gives a financial meaning to the λ as it is obtained through the market data rather than being selected arbitrarily [5, 45]. This gives financial interpretation to λ unlike other regularization methods. The resultant MAP estimate gives an equivalence estimates for the functional minimization of (2.7) explained in Section 2.2.2.

Second, the Bayesian method offers a consistent way of incorporating all known and available information into the prior density, thereby distinguishing between prior information and observed market data. This gives more correctness and meaning to the Bayesian estimators $q_{H_1}(v)$ and $q_{H_2}(v)$. In fact, the Bayesian approach allows a coherent way of attaching weights and confidence level to point estimates making it easier to interpret them.

Third, as mentioned in the introductory section of this chapter, some Bayesian critics who oppose the use of a prior information argue that this methodology often leads to preconceived bias and may dominate the result [11, 40]. They contend that this can cause the Bayesian estimates to reflect more of the prior information than the observed data. However, in the context of this thesis, the parameter we wish to estimate is of financial nature, therefore, the Bayesian approach offers a natural way of incorporating financial assumptions about this parameter. These assumptions have been summarized in Section 2.3.2. It is also important to note that financial concepts such as *no-arbitrage conditions*

and *market completeness* are crucial to pricing and hedging financial instruments. These financial concepts should carefully be incorporated into the financial model. The Bayesian method offers a unique way to do this through its prior density. For example, the no-arbitrage conditions can be incorporated into the prior density by attaching a zero prior probability to parameters that violate the arbitrage conditions for the calibrated instrument.

Furthermore, the critics of Bayesian method of data analysis argue that an unknown model parameter should not be treated as a random variable by constructing a distribution for it [11]. These opponents explain that although the parameter is unknown, it does not have uncertainty. An argument against this is that it is equally important to measure the uncertainty of a calibrated parameter as it is to find the model parameter. Therefore, treating the unknown parameter as a random variable allows for measuring the potential model errors.

Finally, some opponents also argue that assignment of probabilities (weights) to different parameter candidates are too arbitrary, therefore, it is difficult to understand the accuracy of such assignments. However, in the context of financial model studied in this thesis, we argue that it is in fact natural to assign prior weights to different candidates of model parameters. This can be tacitly done by assigning these probabilities via a prior density. Therefore, using a prior density becomes imperative and inevitable as these probability assignments give a form of regularization to the ill-posed problem. Although, authors such as Cont [22] argues that it is quite sophisticated to assign these weights, the Bayesian technique is very robust to inaccurate specifications of the prior weights.

As we have demonstrated in this section, Bayesian methodology has a very robust structure that allows regularization to be done with meaningful interpretation rather than arbitrary assignment of the smoothening parameter. For more details on the Bayesian inference and applications, we refer [11, 40, 46, 82] to the readers.

4.1.3 Literature Review on the Financial Applications of Bayesian method

In the last three decades, there have been an increased amount of research in applying Bayesian learning to calibrate parameters in financial models. Some of the recent literature on the application of Bayeisan method in parameter estimation were mentioned in Section 2.5. Here, we summarize the rest of the unmentioned literature on Bayesian application to financial problems. One of the earliest research on this subject dates back to early 1990s when Jacquier et al. [57] showed that Bayes' estimators for a specific class of stochastic volatility model outperforms moments and quasi-maximum likelihood estimators in hedging. More recently, Animoku [4] showed that Bayesian approach performs better than *frequentist* approach in determining the macroeconomic factors that indicate high level of non-performing loans in the banking sector.

Bhar et al. [12] have used dynamic Bayesian approach in estimating instantaneous spot interest rates. Jacquier & Jarrow [56] used the Black-Scholes model to study the effect of Bayesian estimators on parameter uncertainty and model errors. With their approach, they assessed the non-normality of the posterior distributions by inferring the option price values using Bayesian estimators. Furthermore, Monoyios [71] studied the effects of uncertainty of drift parameter of a traded asset on hedging a correlated non-traded asset in an incomplete market. The author obtained a very large hedging errors. Furthermore, Jobert et al. [58] attempted to explain the consistently excess return observed for a risky asset over T-bills using Bayesian analysis. By replacing the assumptions on dividend parameters with prior a information, the authors showed that the excess rate of return observed is expected.

For Bayesian applications on parameter calibration using local volatility model are the works of Animoku et al. [5] and Gupta & Reisinger [45] whose work have been explained in Section 2.5. There are other works that are closely related to Bayesian study on local volatility as well. For example, Gupta et al. [46] studied the model and parameter uncertainty for local volatility model. Using Bayes' estimates, the model hedging errors are checked for different risk measures. The

authors conclude from their results that derivative pricing is very sensitive to model and parameter uncertainty. In addition, Darsinos & Satchell [25] constructed a posterior distribution for asset price S_t and Black-Scholes European call price C_{BS} by setting a prior density for S_t and Black-Scholes implied volatility σ_{IMP}. The posterior density is then updated whenever there are newly observed prices. Using the posterior density they have constructed, the authors computed the marginal probability distribution for the call option price C_{BS}. Subsequently, in [26], the same authors used the density obtained in [25] to forecast one day ahead European call option prices. Their results showed smaller hedging errors and improved hedging profits when compared to procedures that involve proxies for the implied volatility.

4.2 Consistency of Bayes' Estimators

In this section, we will prove the consistency of Bayes' estimators. Consistency in this context answers the question of whether a posterior density gives more accurate results as more and more observations are observed. In other words, does new knowledge improves the preciseness of the posterior density when updated. As noted by [42, 45], it is crucial to ascertain this property for posterior densities as inconsistent densities can not guarantee inferences made based on them. In this respect, one may expect that as more data values Υ are observed through time, the parameter estimator converges to the true value in some sense. For definitions regarding different forms of convergences, we refer the reader to Appendix A.

Also, consistency is important as prior distribution can sometimes dominate the posterior and possibly lead to inconsistent estimates. Therefore, for call option prices that are independent but not necessarily identically distributed, we prove the consistency of the Bayes' estimators. For the coming sections, we establish the consistency for a single price observation and a scalar parameter, multiple price observations and a scalar parameter; for both additive and multiplicative error models. We further generalize the assertions for the non-scalar parameters.

Definition 4.2 (Consistency)**.** A sequence of estimators $\hat{q}_n$ of an unknown pa-

rameter q over a defined sample space $(\Omega_n, \mathcal{B}_n)$ is said to be *consistent* if it converges to q in probability (written as $\hat{q}_n \xrightarrow{P} q \;\; \forall q \in \Omega$), i.e.,

$$\forall q \in \Omega \;\; \forall \delta > 0 \quad \mathbb{P}_q\left[|\hat{q}_n - q| \geq \delta\right] \to 0 \quad \text{as } n \to \infty, \tag{4.4}$$

where $\mathbb{P}_q$ denotes the probability measure for parameter q.

According to Definition 4.2, as more data is collected over time, the Bayes' estimator of the unknown parameter should get closer to the true parameter value. This is important as inferences made based on inconsistent posterior distribution casts serious doubts on the results [42]. Therefore, it becomes necessary that we prove the consistency of the Bayes' estimators. Here with, we show the consistency under the assumption that the observed data are independent but not necessarily identically distributed.

4.2.1 Literature Review on Consistent Bayes' Estimators

Most of the literature on the consistency of Bayes' estimators assume the observed data are independent and identically distributed (i.i.d.). More recently, [46] has proven for the case where the observations are independent but not necessarily identically distributed. We briefly review some of the findings on the consistency of the Bayesian estimators.

Doob [31] showed the consistency for the Bayes' estimators for every prior distribution $\pi_0(q)$ and a convex loss function, except possibly for q contained in a null sets with respect to the prior. Doob's result in 1949 was one of the earliest breakthroughs in studying consistent Bayes' estimators. Consequently, in 1965 Schwartz [79] extended the work in [31] to non-convex loss functions. In a more recent work, Fitzpatrick[38] has derived the consistency for the *maximum likelihood estimators* in the case of Gaussian noise.

There are also some extensive research on the inconsistency of Bayesian estimators for some choices of prior. For example, Diaconis & Freedman [30] showed that the use of Dirichlet prior can lead to inconsistent Bayes' estimates. Furthermore, Wasserman [86] showed that for some posterior density with infinite

dimensional parameter, the Bayes' estimators are inconsistent and suggested a *frequentist* approach instead.

Further research has been done by Le Cam [66] in establishing the conditions for a Bayes' estimator to be consistent given a suitable prior distribution. The author proved that under certain conditions the consistency of the maximum likelihood estimator implies the consistency of the Bayes' estimator. Strasser [83] improved the work of [66] by proving that every condition that implies consistency for the maximum likelihood estimator also implies the consistency of the Bayes' estimator for a large class of prior distributions. Furthermore, in [13], Bickel & Yahav showed that the results in [66] can be applied to a larger class of loss functions.

Barron et al. [9] proved the consistency of Bayes' estimators by showing that as more data is observed, the probability of the posterior density lying within an ϵ-neighborhood of the true density tends to 1. As an extension, authors such as Shen & Wasserman [80] and Ghosal et al. [43], have investigated the rates of convergence of the posterior distributions with respect to the true parameter for non i.i.d. observations.

All the literature discussed so far are research done for the case of i.i.d. observations. In a more recent work, Gupta & Reisinger [45] have shown the consistency for the Bayes' estimators (for additive normally distributed error model) where observations are independent but not identically distributed. We have summarized the key results from [45] with some adjustments to their proofs in the next couple of sections. Furthermore, we have improved the theoretical results obtained in [45] by proving the consistency of Bayes' estimators for multiplicative normally distributed error model.

4.2.2 Consistency of Bayes' Estimator for Additive Error Model: Single Price Observation and Scalar Parameter

Consider the following statistical model

$$V_t = f_t(\sigma) + e_t, \tag{4.5}$$

where V_t is the market option price; $f_t(\sigma_{LV})$ is the model response from the local volatility model parametrized by σ and $e_t \sim \mathcal{N}(0, \epsilon_t^2)$, the market noise. Suppose we have only one volatility parameter to estimate from the local volatility model. Also, let's assume we have one observation of the option price $V_t^{(1)}(\sigma)$ at each time t. At time t_n, we have a total of n observations. Further assume that the market noise at each time period is distributed as

$$e_t^{(1)} = \mathcal{N}(0, (\epsilon_t^{(1)})^2)$$

for $\epsilon_t^{(1)} > 0$, $\forall t \in \{t_1, t_2, \ldots, t_n\}$ that are independent but not necessarily identically distributed.

Definition 4.3. Define

$$\mathcal{G}_{t_n} = \sigma(\{e_s : s \in \Upsilon\})$$

to be the sigma algebra generated by the errors where

$$\Upsilon_n = \Upsilon_n([0, T]) = \{t_i, i = 1, \ldots, n : 0 = t_1 < t_2 < \ldots < t_n = T\}.$$

Then, assuming the independence of the price process and random error process, i.e $\mathcal{F}_{t_n} \perp \mathcal{G}_{t_m} \ \forall \ (n, m)$, we have the posterior density using Bayes' formula as:

$$\begin{aligned}
\pi(\sigma_n(V)) = \pi_n(\sigma|V) &= \frac{\pi_n(V|\sigma)\pi_0(\sigma)}{\pi_n(V)} \\
&= \frac{\pi(V_{t_1}|\sigma) \cdots \pi(V_{t_n}|\sigma)\pi_0(\sigma)}{\pi_n(V)} \\
&= \prod_{t \in \Upsilon_n} \frac{1}{\sqrt{2\pi}\epsilon_t} \exp\left\{-\frac{1}{2\epsilon_t^2}(V_t - f_t(\sigma)^2)\right\} \frac{\pi_0(\sigma)}{\pi_n(V)},
\end{aligned}$$

where $f_t(\sigma)$ is the model process from solving the local volatility equation.

Given the loss function $H(\sigma, \sigma')$, we define the sequence of Bayes' estimators $\hat{\sigma}$ by

$$\hat{\sigma}_n(V) = \underset{\sigma' \in \mathcal{Q}}{\operatorname{argmin}} \{g(\sigma_n(V), \sigma')\}, \qquad (4.6)$$

where

$$g(\sigma_n(V), \sigma') = \mathbb{E}[H(\sigma_n(V), \sigma')] = \int_\sigma H(\sigma_n(V), \sigma')\pi_n(\sigma|V)d\sigma,$$

$\mathcal{Q}$ is the support for the parameter σ under the prior density $\pi_0(\sigma)$, and $\pi_n(V)$ is the posterior density described by (4.2). Note that $\hat{\sigma}_n(V)$ is not necessarily

unique. We are interested in showing that the estimator $\hat{\sigma}_n(V)$ converges in probability to some true value σ^*.

Assumption 4.2 ([45], Assumption 2.3). *The prior density $\pi_0(\sigma)$ and its support $\mathcal{Q}$ satisfy:*

- $\pi_0(\sigma)$ *is continuous at σ^*;*

- $\pi_0(\sigma) > 0;$

- $\pi_0(\sigma)$ *is finite on $\mathcal{Q}$;*

- $\mathcal{Q}$ *is compact.*

Assumption 4.3 ([45], Assumption 2.4). *There exists a finite $k > 0$ such that for all σ, $\sigma^* \in \mathcal{Q}$, $|\sigma - \sigma^*| > 0$, for every t, the functions $f_t(\sigma)$ and noise standard deviations ϵ_t conditional on $\mathcal{F}_t$ and $\mathcal{G}_t$ respectively, satisfy*

$$\frac{1}{\epsilon_t} \frac{|f_t(\sigma) - f_t(\sigma^*)|}{|\sigma - \sigma^*|} \geq k.$$

Note that Assumption 4.2 gives the properties of the prior density and specifically the fourth property ensures that the parameter space is closed and bounded. Furthermore, Assumption 4.3 implies that all the f_t are strictly monotone, which is reasonable since Black-Scholes model satisfies this property for European call price functions. Also, ϵ_t, the noise standard deviation is expected to be finite in the case of Black Scholes model.

Lemma 4.4 ([45], Lemma 2.5). $\sigma_n(V) = \sigma(V_{t_n}) \xrightarrow{P} \sigma^*$ *as $n \to \infty$.*

Proof. Consider a fixed f_t, $\forall t \in \Upsilon_n$. This means that the stock price process is fixed up to time t_n and all prices are $\mathcal{F}_t$ measurable. Therefore, P refers to only the randomness caused by the market noise identified as e_t. So, we can write the posterior distribution at time t_n as

$$\pi(\sigma|V) = \gamma_n(V)e^{-\frac{1}{2}\phi_n(\sigma,V)},$$

50

where $\gamma_n(V)$ is the normalization constant with respect to σ and

$$\phi_n(\sigma, V) = -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \frac{1}{\epsilon_t^2}(V_t - f_t(\sigma))^2$$

$$= -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \left(\frac{V_t - f_t(\sigma^*)}{\epsilon_t} - \frac{f_t(\sigma) - f_t(\sigma^*)}{\epsilon_t} \right)^2$$

$$= -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \left(Z_t - \frac{f_t(\sigma) - f_t(\sigma^*)}{\epsilon_t} \right)^2.$$

Z_t are taken to be independent standard Gaussian random variables. Using the first and second properties of Assumption 4.2, there exists $\eta_0 > 0$ such that for every $\sigma \in [\sigma^* - \eta_0, \sigma^* + \eta_0]$, $\pi_0(\sigma) \geq d > 0$. Also, by the fourth property of Assumption 4.2, we specify the support $\mathcal{Q} \subset [\underline{\sigma}, \overline{\sigma}]$ for the posterior density for some constants $\underline{\sigma}, \overline{\sigma} \in \mathbb{R}$.

We aim to show that the probability distribution function over the support of the parameter σ accumulates its mass at σ^* as $n \to \infty$. That is, we need to show

$$\int_{\underline{\sigma}}^{\overline{\sigma}} \pi_n(\sigma|V)d\sigma = \int_{\sigma^* - \eta}^{\sigma^* + \eta} \pi_n(\sigma|V)d\sigma = 1 \tag{4.7}$$

for $\eta_0 > \eta > 0$ as $n \to \infty$.

First, we will establish that

$$\int_{\underline{\sigma}}^{\sigma^* - \eta} \pi_n(\sigma|V)d\sigma = \int_{\sigma^* + \eta}^{\overline{\sigma}} \pi_n(\sigma|V)d\sigma = 0 \tag{4.8}$$

as $n \to \infty$. Now, for all $\eta > 0$ such that $\eta_0 > \eta$,

$$\int_{\underline{\sigma}}^{(\sigma^* - \eta)} \pi_n(\sigma|V)d\sigma = \int_{\underline{\sigma}}^{(\sigma^* - \eta)} \gamma_n(V)e^{-\frac{1}{2}\phi_n(\sigma, V)}d\sigma$$

$$\leq [(\sigma^* - \eta) - \underline{\sigma}]\, \gamma_n(V)e^{-\frac{1}{2}\phi_n(\alpha_n, V)}, \tag{4.9}$$

where $\alpha_n = \mathrm{argmax}_{[\underline{\sigma}, \sigma^* - \eta]}\, e^{-\frac{1}{2}\phi_n(\sigma, V)}$. Also, observe that

$$\gamma_n(V) = \frac{1}{\int_{\underline{\sigma}}^{(\sigma^* - \eta)} e^{-\frac{1}{2}\phi_n(\sigma, V)}d\sigma} \leq \frac{1}{[(\sigma^* - \eta) - \underline{\sigma}]}e^{\frac{1}{2}\phi_n(\beta_n, V)},$$

with $\beta_n = \mathrm{argmin}_{[\underline{\sigma}, \sigma^* - \eta]}\, e^{-\frac{1}{2}\phi_n(\sigma, V)}$. Thus, substituting α_n and β_n into (4.9) yields

$$\int_{\underline{\sigma}}^{(\sigma^* - \eta)} \pi_n(\sigma|V)d\sigma \leq e^{-\frac{1}{2}[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]},$$

51

where

$$\phi_n(\alpha_n, V) - \phi_n(\beta_n, V) = -2\log(\pi_0(\alpha_n)) + 2\log(\pi_0(\beta_n))$$
$$+ \sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]. \quad (4.10)$$

By the third property of Assumption 4.2, $\pi_0(\beta_n), \pi_0(\alpha_n) < \infty$. Therefore, we can bound the first two terms of (4.10) by some constant, that is,

$$-2\log(\pi_0(\alpha_n)) + 2\log(\pi_0(\beta_n)) > -\infty.$$

Now, expanding the remaining terms in (4.10), and using $\beta_n < \alpha_n \leq \sigma^*$ with f_t strictly increasing give

$$\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]$$
$$= 2\sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) + \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]$$
$$\geq 2\sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) + \sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2 = X_n,$$

$$\quad (4.11)$$

where X_n is a normally distributed with mean

$$\mathbb{E}[X_n] = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2 + 2\sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) \mathbb{E}[Z_t]$$
$$= \sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2,$$

and variance

$$\mathbb{V}[X_n] = 4\sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2 \mathbb{V}[Z_t]$$
$$= 4\sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2.$$

Here, $Z_t \sim \mathcal{N}(0,1)$ for all $t \in \Upsilon_n$. Thus, $X_n \sim \mathcal{N}(s_n, 4s_n)$, with

$$s_n = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2 = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t} \right)^2.$$

Furthermore, note that $|\alpha_n - \beta_n| \geq \epsilon > 0$ by using the definitions of α_n and β_n. Assumption 4.3 implies that

$$s_n \geq \sum_{t \in \Upsilon_n} k^2 |\alpha_n - \beta_n|^2 \geq nk^2\epsilon^2 \to +\infty.$$

Also, observe that for any $r \in \mathbb{R}^+$, the probability

$$
\begin{aligned}
\mathbb{P}[e^{-\frac{1}{2}X_n} \geq r] = \mathbb{P}[X_n \leq -2\log r] &= \int_{-\infty}^{-2\log r} \frac{1}{\sqrt{8\pi s_n}} e^{-\frac{1}{8s_n}(x-s_n)^2} dx \\
&= \int_{-\infty}^{\frac{-2\log r - s_n}{\sqrt{4s_n}}} \frac{1}{\sqrt{8\pi s_n}} e^{-\frac{1}{8s_n}(y\sqrt{4s_n})^2} \sqrt{4s_n} dy \\
&= \int_{-\infty}^{\frac{-2\log r - s_n}{\sqrt{4s_n}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\
&= \Phi\left(\frac{-2\log r - s_n}{\sqrt{4s_n}}\right) \\
&\to \Phi(-\infty) \\
&= 0 \quad \text{as} \quad n \to \infty,
\end{aligned}
$$

where Φ is the cumulative distribution function for a standard normal random variable. Thus, $e^{-\frac{1}{2}X_n} \xrightarrow{P} 0$ and

$$0 \leq \int_{\underline{\sigma}}^{(\sigma^*-\eta)} \pi_n(\sigma|V) d\sigma \leq e^{-\frac{1}{2}[\phi_n(\alpha_n,V)-\phi_n(\beta_n,V)]} \leq e^{-\frac{1}{2}X_n} \xrightarrow{P} 0 \qquad (4.12)$$

as $n \to \infty$. By *squeeze theorem*, we conclude (4.8).

By a similar line of reasoning we can also show that

$$\int_{\sigma^*+\eta}^{\overline{\sigma}} \pi_n(\sigma|V) d\sigma = 0,$$

for $\eta_0 > \eta > 0$. We conclude that the assertion in (4.7) is true and $\sigma_n(V) = \sigma(V_{t_n}) \xrightarrow{P} \sigma^*$ as $n \to \infty$. $\qquad\square$

Theorem 4.5 ([45], Theorem 2.6). *For a function H that is bounded and continuous on $\mathcal{Q}$, the Bayes' estimator $\hat{\sigma}_n(V)$ is consistent.*

Proof. Observe that we can write $H(\sigma, \sigma') = h_\sigma(\sigma - \sigma')$ for a fixed σ and some function $h_\sigma : \mathbb{R} \to \mathbb{R}$. For $\delta > 0$,

$$
\begin{aligned}
\mathbb{P}_{\sigma^*}\left[|\hat{\sigma}_n(V) - \sigma^*| \geq \delta\right] &\leq \mathbb{P}_{\sigma^*}\left[|\hat{\sigma}_n(V) - \sigma_n(V)| + |\sigma_n(V) - \sigma^*| \geq \delta\right] \\
&\leq \mathbb{P}_{\sigma^*}\left[|\hat{\sigma}_n(V) - \sigma_n(V)| \geq \frac{1}{2}\delta\right] + \mathbb{P}_{\sigma^*}\left[|\sigma_n(V) - \sigma^*| \geq \frac{1}{2}\delta\right],
\end{aligned}
$$

where $\mathbb{P}_{\sigma^*}$ represents the probability with respect to the model with parameter σ^*. We know from Lemma 4.4 that $\sigma_n(V) \overset{P}{\to} \sigma^*$, which implies that $\mathbb{P}_{\sigma^*}\left[|\sigma_n(V) - \sigma^*| \geq \frac{1}{2}\delta\right] \to 0$ as $n \to \infty$. Also,

$$\mathbb{P}_{\sigma^*}\left[|\hat{\sigma}_n(V) - \sigma_n(V)| \geq \frac{1}{2}\delta\right]$$

$$\leq \mathbb{P}_{\sigma^*}\left[H\left(\hat{\sigma}_n(V), \sigma_n(V)\right) \geq \min\left\{h_{\hat{\sigma}_n(V)}(\frac{1}{2}\delta), h_{\hat{\sigma}_n(V)}(-\frac{1}{2}\delta)\right\}\right]$$

$$\leq \frac{\mathbb{E}_{\sigma^*}\left[H\left(\hat{\sigma}_n(V), \sigma_n(V)\right)\right]}{\min\left\{h_{\hat{\sigma}_n(V)}(\frac{1}{2}\delta), h_{\hat{\sigma}_n(V)}(-\frac{1}{2}\delta)\right\}} \quad \text{(Use of markov property)}$$

$$\leq \frac{\mathbb{E}_{\sigma^*}\left[H\left(\hat{\sigma}_n(V), \sigma^*\right)\right]}{\min\left\{h_{\hat{\sigma}_n(V)}(\frac{1}{2}\delta), h_{\hat{\sigma}_n(V)}(-\frac{1}{2}\delta)\right\}} \to 0 \quad \text{(By minimality of } \hat{\sigma}_n(V))$$

as $n \to \infty$. In the last line we have used the fact that H is bounded and continuous with $\sigma_n(V) \overset{D}{\to} \sigma^*$. as $n \to \infty$. We conclude that the Bayes' estimator is consistent since for $\delta > 0$, $\mathbb{P}_{\sigma^*}\left[|\hat{\sigma}_n(V) - \sigma^*| \geq \delta\right] \to 0$ as $n \to \infty$. $\qquad \square$

Remark 4.2. For the case where we have several price observations $V_t^{(i)}$ at each time t, that is $|I_t| > 1$, we have similar conclusions as in Lemma 4.4 and Theorem 4.5. The proof is analogous giving the assumption that market noises are independent and normally distributed as before, including the prices corresponding to the same time t. We follow this with a corollary.

Corollary 4.6. *For multiple price observation set $|I_t| > 1$ and all loss functions H that is bounded and continuous on $\mathcal{Q}$, the Bayes' estimator $\hat{\sigma}(V)$ is consistent.*

Proof. The proof is analogous to those of Lemma 4.4 and Theorem 4.5 except that the sigma-field generated by all market errors up to time t_n is now

$$\mathcal{G}_{t_n} = \sigma\left(\{e_s^{(i)} : s \in \Upsilon_n, i \in I_t\}\right)$$

and all the rest of the notations should be replaced as follows:

$$\sum_{t \in \Upsilon_n} \text{ by } \sum_{t \in \Upsilon_n, i \in I_t}, \quad V_t \text{ by } V_t^{(i)}, \quad f_t \text{ by } f_t^{(i)}, \quad \epsilon_t \text{ by } \epsilon_t^{(i)}.$$

$\qquad \square$

4.2.3 Consistency of Bayes' Estimator for Additive Error Model: Non-Scalar Parameter

Here, we consider the parameter $\sigma \in \mathbb{R}^p$ to be finite dimensional but non-scalar with dimension p. Suppose the norm $\|\cdot\|$ associated with the parameter space, $\mathbb{R}^p$, is finite. We extend the results of the previous sections to the case of non-scalar parameter. The assumptions in Section 4.2.2 hold with some modifications to Assumption 4.3.

Assumption 4.7 ([45], Assumption 2.10). *For each $\sigma \in \mathcal{Q}$, $\exists n(\sigma) \in \mathbb{N}$ and $K, k \in \mathbb{R}^+$ such that*

$$K^2 > \frac{1}{n} \sum_{t \in \Upsilon_n} \frac{1}{\epsilon_t^2} \frac{|f_t(\sigma) - f_t(\sigma^*)|^2}{\|\sigma - \sigma^*\|^2} > k^2, \quad \forall n \geq n(\sigma).$$

Assumption 4.7 simply states that the average computed squared error is bounded by non-zero constants. For the scalar case $n(\sigma) = 1$, $\forall \sigma \in \mathcal{Q}$.

Theorem 4.8 ([45], Theorem 2.6). *For all H bounded and continuous functions on σ, the non- scalar Bayes' estimator $\hat{\sigma}_n(V)$ is consistent under the additive normally distributed error model.*

Proof. The proof is analogous to the proof of Lemma 4.4 with some adjustments for the higher dimensionality of σ. Here with, instead of integrals on real line, we take the integrals in p-dimensional space. So, we need to show the probability distribution function $\pi_n(\sigma|V)$ over the support of the parameter σ accumulates its mass on a ball in a p-dimensional space centered around σ^*, as $n \to \infty$. Mathematically, we should show that

$$\int_{B(\sigma^*,\eta)^c} \pi_n(\sigma|V)d\sigma = 0 \ \text{ as } \ n \to \infty, \tag{4.13}$$

for all $\eta_0 > \eta > 0$ such that $\|\sigma - \sigma^*\| \leq \eta_0$, and $\pi_0(\sigma) \geq d > 0$. Here, 'c' denotes the complement of a set in $\mathbb{R}^p$. Again, by similar computation in the proof of Lemma 4.4, we can bound (4.13) by

$$\int_{B(\sigma^*,\eta)^c} \pi_n(\sigma|V)d\sigma \leq e^{-\frac{1}{2}[\phi_n(\alpha_n,V) - \phi_n(\beta_n,V)]},$$

where $\alpha_n, \beta_n \in \mathcal{Q}$ are defined similarly as

$$\alpha_n = \operatorname*{argmax}_{B(\sigma^*, \eta)^c} e^{-\frac{1}{2}\phi_n(\sigma, V)} \quad \text{and} \quad \beta_n = \operatorname*{argmin}_{B(\sigma^*, \eta)^c} e^{-\frac{1}{2}\phi_n(\sigma, V)}.$$

Again, writing the expression for $\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)$ and using the fact that $\pi_0(\beta_n), \pi_0(\alpha_n) < \infty$ by the third property of Assumption 4.2 together with $n \geq n(\alpha_n), n(\beta_n)$ yields

$$\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]$$

$$= 2 \sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) + \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]$$

$$\geq 2 \sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) + n\eta_1^2(k^2 - K^2) = X_n, \quad \text{(Using Assumption 4.7)}$$

$$\tag{4.14}$$

where $\eta_1^2 = \|\alpha_n - \sigma^*\|^2$ and $X_n \sim \mathcal{N}(n\eta_1^2(k^2 - K^2), 4s_n)$ is a random variable; s_n is bounded by

$$s_n = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2 = \sum_{t \in \Upsilon_n} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t} \right)^2$$

$$\geq \sum_{t \in \Upsilon_n} k^2 \|\alpha_n - \beta_n\|^2 \geq nk^2\epsilon^2 \longrightarrow +\infty.$$

We have used Assumption 4.7 in the second line. Hence, we can also show as before that for any $r \in \mathbb{R}^+$, $\mathbb{P}[e^{-\frac{1}{2}X_n} \geq r] \to 0$. The rest of the arguments follows from the remaining proofs of Lemma 4.4 and Theorem 4.5. $\qquad \square$

Example 4.1. Let's consider the Black-Scholes model and its framework where σ is the volatility parameter associated with the model. The observed option prices under this model are analytically given as

$$f_t(\sigma) = e^{-r(T-t)} \mathbb{E}\left[(S_T(\sigma) - K)^+ | \mathcal{F}_t \right].$$

We know that the f_t is differentiable everywhere on $[\sigma_{min}, \sigma_{max}]$ for $0 < \sigma_{min} < \sigma_{max} < \infty$, and the vega, which is the derivative of the option price with respect to the volatility, is given by

$$f_t'(\sigma) = S_t \sqrt{T-t} N' \left(\frac{\log(S_t/K) + (r + \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}} \right),$$

56

for a non-dividend paying stock S where $N'(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. The vega is positive for all combinations of S_t and K, making f_t a monotone increasing function bounded below by $f_t(\sigma_{min})$ on the interval $[\sigma_{min}, \sigma_{max}]$. Thus, assuming finite market noise, the Black-Scholes European call option prices satisfy Assumption 4.3. So, we can justifiably construct consistent Bayes' estimators for the volatility using the theoretical framework outlined in this thesis.

4.2.4 Consistency of Bayes' Estimator for Multiplicative Error Model: Single Price Observation and Scalar Parameter

Consider the settings in Section 4.2.2 where we have a single price observation at each time $t \in \Upsilon_n$ and a single parameter (in this case volatility) to be estimated. Furthermore, replace the statistical model (additive normally distributed error) considered in Section 4.2.2 with the multiplicative normally distributed error model below:

$$V_t = f_t(\sigma_{LV})\,(1 + e_t),$$

where the market noise e_t is distributed as before. Unless stated otherwise, the set of assumptions and definitions used in Section 4.2.2 are also applied here. Thus, the posterior density over the n period can be written as

$$
\begin{aligned}
\pi(\sigma_n(V)) = \pi_n(\sigma|V) &= \frac{\pi_n(V|\sigma)\pi_0(\sigma)}{\pi_n(V)} \\
&= \frac{\pi(V_{t_1}|\sigma)\cdots\pi(V_{t_n}|\sigma)\pi_0(\sigma)}{\pi_n(V)} \\
&= \prod_{t \in \Upsilon_n} \frac{1}{\sqrt{2\pi}e_t} \exp\left\{-\frac{1}{2e_t^2 f_t(\sigma)^2}(V_t - f_t(\sigma)^2)\right\} \frac{\pi_0(\sigma)}{\pi_n(V)}.
\end{aligned}
$$

$$(4.15)$$

Lemma 4.9. $\sigma_n(V) = \sigma(V_{t_n}) \xrightarrow{P} \sigma^*$ *as $n \to \infty$ for the multiplicative normally distributed error model.*

Proof. Again, we consider a fixed f_t, for all $t \in \Upsilon_n$ such that the only randomness is coming from the market noise e_t. We can then write the posterior density at time t_n as

$$\pi(\sigma|V) = \gamma_n(V)e^{-\frac{1}{2}\phi_n(\sigma, V)},$$

where $\gamma_n(V)$ is the normalization constant with respect to σ and $\phi_n(\sigma, V)$ given by

$$\phi_n(\sigma, V) = -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \frac{1}{\epsilon_t^2 f_t(\sigma)}(V_t - f_t(\sigma))^2$$

$$= -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \left(\frac{V_t - f_t(\sigma^*)}{\epsilon_t f_t(\sigma)} - \frac{f_t(\sigma) - f_t(\sigma^*)}{\epsilon_t f_t(\sigma)} \right)^2$$

$$= -2\log(\pi_0(\sigma)) + \sum_{t \in \Upsilon_n} \left(Z_t - \frac{f_t(\sigma) - f_t(\sigma^*)}{\epsilon_t} f_t(\sigma) \right)^2.$$

Z_t is taken to be independent standard Gaussian random variables.

We aim to show that the probability distribution function over the support of the parameter σ accumulates its mass at σ^* as $n \to \infty$. That is, we need to establish that

$$\int_{\underline{\sigma}}^{\sigma^* - \eta} \pi_n(\sigma|V) d\sigma = \int_{\sigma^* + \eta}^{\overline{\sigma}} \pi_n(\sigma|V) d\sigma = 0, \tag{4.16}$$

as $n \to \infty$. Again, by similar computation as shown in the proof of Lemma 4.4, we can bound (4.16) by

$$\int_{\underline{\sigma}}^{(\sigma^* - \eta)} \pi_n(\sigma|V) d\sigma \leq e^{-\frac{1}{2}[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]},$$

where $\alpha_n, \beta_n \in \mathcal{Q}$ are defined similarly as

$$\alpha_n = \underset{[\underline{\sigma}, \sigma^* - \eta]}{\operatorname{argmax}} e^{-\frac{1}{2}\phi_n(\sigma, V)} \quad \text{and} \quad \beta_n = \underset{[\underline{\sigma}, \sigma^* - \eta]}{\operatorname{argmin}} e^{-\frac{1}{2}\phi_n(\sigma, V)}.$$

and

$$\phi_n(\alpha_n, V) - \phi_n(\beta_n, V) = -2\log(\pi_0(\alpha_n)) + 2\log(\pi_0(\beta_n)) +$$

$$\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right]. \tag{4.17}$$

Again, by bounding the first two terms and using $\beta_n < \alpha_n \leq \sigma^*$ with f_t strictly

increasing, we can expand (4.17) by

$$
\begin{aligned}
\sum_{t \in \Upsilon_n} &\left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right] \\
&= 2 \sum_{t \in \Upsilon_n} \frac{Z_t}{\epsilon_t} \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{f_t(\beta_n)} - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{f_t(\alpha_n)} \right) \\
&\quad + \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right] \\
&\geq -2 \sum_{t \in \Upsilon_n} Z_t \left[\frac{f_t(\sigma^*)}{\epsilon_t} \left(\frac{1}{f_t(\beta_n)} - \frac{1}{f_t(\alpha_n)} \right) \right] \\
&\quad + \sum_{t \in \Upsilon_n} \left[\frac{f_t(\sigma^*)^2}{\epsilon_t^2} \left(\frac{1}{f_t(\beta_n)^2} - \frac{1}{f_t(\alpha_n)^2} \right) \right] \\
&= Y_n,
\end{aligned}
\tag{4.18}
$$

where $Y_n \sim \mathcal{N}(r_n, 4m_n)$ is a random variable with

$$
r_n = \sum_{t \in \Upsilon_n} \left[\frac{f_t(\sigma^*)^2}{\epsilon_t^2} \left(\frac{1}{f_t(\beta_n)^2} - \frac{1}{f_t(\alpha_n)^2} \right) \right],
$$

and

$$
m_n = \sum_{t \in \Upsilon_n} \left[\frac{f_t(\sigma^*)}{\epsilon_t} \left(\frac{1}{f_t(\beta_n)} - \frac{1}{f_t(\alpha_n)} \right) \right]^2.
$$

Now, observe that since $f_t(\sigma^*) \geq f_t(\alpha_n)$,

$$
\begin{aligned}
m_n &= \sum_{t \in \Upsilon_n} \left[\frac{f_t(\sigma^*)}{\epsilon_t} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{f_t(\alpha_n) f_t(\beta_n)} \right) \right]^2 \\
&\geq \sum_{t \in \Upsilon_n} \left[\frac{1}{\epsilon_t} \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{f_t(\beta_n)} \right) \right]^2 \\
&\geq \sum_{t \in \Upsilon_n} \left[\frac{1}{\epsilon_t} \left(\frac{f_t(\alpha_n)}{f_t(\beta_n)} - 1 \right) \right]^2 \\
&> nk^2 \to \infty \quad \text{as} \quad n \to \infty.
\end{aligned}
\tag{4.19}
$$

In line four we have used the fact that

$$
\frac{1}{\epsilon_t} \left(\frac{f_t(\alpha_n)}{f_t(\beta_n)} - 1 \right) > k
$$

for constant $k \in \mathbb{R}^+$. For $\beta_n < \alpha_n$, we have that $\frac{f_t(\alpha_n)}{f_t(\beta_n)} - 1 > \zeta > 0$. Since the market noise ϵ_t is finite, we can find such k that satisfies

$$
\frac{1}{\epsilon_t} \left(\frac{f_t(\alpha_n)}{f_t(\beta_n)} - 1 \right) > \frac{\zeta}{\epsilon_t} = k.
$$

Furthermore, we can also show as before that for any $r \in \mathbb{R}^+$, $\mathbb{P}[e^{-\frac{1}{2}X_n} \geq r] \to 0$. The rest of the arguments follow from the remaining proofs of Lemma 4.9. $\quad \square$

Theorem 4.10. *For all H bounded and continuous functions on σ, the scalar Bayes' estimator $\hat{\sigma}_n(V)$ is consistent under the multiplicative normally distributed error model.*

Proof. The proof is analogous to the proof of Theorem 4.10. $\square$

Remark 4.3. For the case where we have several price observations $V_t^{(i)}$ at each time t, that is $|I_t| > 1$, we have similar conclusions as in Lemma 4.9 and Theorem 4.10. The proof is analogous given the assumption that the market noises are independent and normally distributed as before, including the prices corresponding to the same time t. We follow this with a corollary.

Corollary 4.11. *For multiple price observation, set $|I_t| > 1$ and for all loss functions H that are bounded and continuous on $\mathcal{Q}$, the Bayes' estimator $\hat{\sigma}(V)$ is consistent.*

Proof. The proof is analogous to those of Lemma 4.9 and Theorem 4.10 except that the sigma-field generated by all market errors up to time t_n is now

$$\mathcal{G}_{t_n} = \sigma\left(\{e_s^{(i)} : s \in \Upsilon_n, i \in I_t\}\right).$$

The rest of the notations should be replaced as follows:

$$\sum_{t \in \Upsilon_n} \text{ by } \sum_{t \in \Upsilon_n, i \in I_t}, \quad V_t \text{ by } V_t^{(i)}, \quad f_t \text{ by } f_t^{(i)}, \quad \epsilon_t \text{ by } \epsilon_t^{(i)}$$

in order to complete the proof. $\square$

4.2.5 Consistency of Bayes' Estimator for Multiplicative Error Model: Non-Scalar Parameter

Consider the settings in Section 4.2.3 where we take the parameter $\sigma \in \mathbb{R}^p$ to be finite dimensional but non-scalar. We extend the results of that section to the case of multiplicative normally distributed error model described in Section 4.2.4.

Theorem 4.12. *For all H bounded and continuous functions on σ, the non-scalar Bayes' estimator $\hat{\sigma}_n(V)$ is consistent under the multiplicative normally distributed error model.*

Proof. The proof is analogous to the proof of Theorem 4.8 with some adjustments. Using the same notation again, we should show that

$$\int_{B(\sigma^*,\eta)^c} \pi_n(\sigma|V)d\sigma = 0 \ \text{ as } \ n \to \infty, \tag{4.20}$$

for all $\eta_0 > \eta > 0$ such that $\|\sigma - \sigma^*\| \leq \eta_0$ and $\pi_0(\sigma) \geq d > 0$. Again, 'c' denotes the complement of a set in $\mathbb{R}^p$. Note that $\pi_n(\sigma|V)$ has been defined in (4.15). Again, by similar computation as in the proof of Lemma 4.9, we can bound (4.20) by

$$\int_{B(\sigma^*,\eta)^c} \pi_n(\sigma|V)d\sigma \leq e^{-\frac{1}{2}[\phi_n(\alpha_n,V) - \phi_n(\beta_n,V)]},$$

where $\alpha_n, \beta_n \in \mathcal{Q}$ are defined similarly as

$$\alpha_n = \operatorname*{argmax}_{B(\sigma^*,\eta)^c} e^{-\frac{1}{2}\phi_n(\sigma,V)} \quad \text{and} \quad \beta_n = \operatorname*{argmin}_{B(\sigma^*,\eta)^c} e^{-\frac{1}{2}\phi_n(\sigma,V)}.$$

Again, writing the expression for $\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)$ and using the fact that $\pi_0(\beta_n), \pi_0(\alpha_n) < \infty$ by the third property of Assumption 4.2 together with $n \geq n(\alpha_n), n(\beta_n)$ yields

$$
\begin{aligned}
\sum_{t\in\Upsilon_n} & \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t}\right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t}\right)^2\right] \\
& = 2\sum_{t\in\Upsilon_n} \frac{Z_t}{\epsilon_t}\left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{f_t(\beta_n)} - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{f_t(\alpha_n)}\right) \\
& \quad + \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)}\right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)}\right)^2\right] \\
& \geq -2\sum_{t\in\Upsilon_n} Z_t\left[\frac{f_t(\sigma^*)}{\epsilon_t}\left(\frac{1}{f_t(\beta_n)} - \frac{1}{f_t(\alpha_n)}\right)\right] \\
& \quad + \sum_{t\in\Upsilon_n} \eta_1^2\left(\frac{k^2}{f_t(\alpha_n)^2} - \frac{K^2}{f_t(\beta_n)^2}\right) \\
& = Y_n,
\end{aligned}
\tag{4.21}
$$

where $\eta_1^2 = \|\alpha_n - \sigma^*\|^2$ and $Y_n \sim \mathcal{N}(r_n, 4m_n)$ is a random variable with m_n bounded by

$$
\begin{aligned}
m_n & = \sum_{t\in\Upsilon_n}\left[\frac{f_t(\sigma^*)}{\epsilon_t}\left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{f_t(\alpha_n)f_t(\beta_n)}\right)\right]^2 \\
& \geq \sum_{t\in\Upsilon_n} k^2\epsilon^2\frac{1}{f_t(\beta_n)^2} \quad \text{(by Assumption 4.7, and } f_t(\sigma^*) \geq f_t(\alpha_n)) \\
& \geq nk^2\epsilon^2\frac{1}{f_t(\beta_n)^2} \to \infty \ \text{ as } \ n \to \infty,
\end{aligned}
$$

61

where $\|\alpha_n - \beta_n\|^2 = \epsilon > 0$, and

$$r_n = \sum_{t \in \Upsilon_n} \eta_1^2 \left(\frac{k^2}{f_t(\alpha_n)^2} - \frac{K^2}{f_t(\beta_n)^2} \right).$$

We have used Assumption 4.7 in the fourth line of (4.21).

Hence, we can also show as before that for any $r \in \mathbb{R}^+$, $\mathbb{P}[e^{-\frac{1}{2}X_n} \geq r] \to 0$. The rest of the arguments follow from the remaining proofs of Lemma 4.9 and Theorem 4.10. $\qquad \square$

4.3 Markov Chain and Markov Chain Monte Carlo (MCMC) Methods

We have discussed so far how to obtain the posterior distribution under Bayesian framework and the Bayes' estimators based on a loss function attached to the posterior density. However, we have not specified the numerical techniques used in sampling the parameter values from the posterior density. In this section, we study the theoretical framework on how to do so. In most practical applications, physical systems are often modeled as Markov chains. This is because most stochastic systems often exhibit the *Markov property*, such that the prediction of the future state is only dependent on its present state. This property of stochastic processes was first studied in the 1900s by Andrey Markov[67].

Herewith, we regard the sampling of the stochastic parameter (local volatility) from its posterior density as a Markov process. By using Markov Chain Monte Carlo (MCMC) technique, we aim to simulate the realizations of this process. Furthermore, to simulate the chain via MCMC, we have used the Metropolis Hasting Algorithm.

The rest of this section is arranged as follows: First, we develop the fundamental theory and properties of Markov chains as well as the detailed balance equation in Section 4.3.1. In Section 4.3.2, we discuss the importance and motivation behind the use of Markov Chain Monte Carlo technique in sampling parameter values from the posterior density. Consequently, in Section 4.3.3, we explain the Metropolis Hasting Algorithm used in the construction of the Markov chains.

Finally, Section 4.3.4 gives analytical result on the convergence criteria for the MCMC chains.

4.3.1 Markov Chains

A stochastic process X_n satisfies a Markov property if the probability of the future states is independent of its history except the current state. In other words, a Markov process has a memoryless property such that past events are not required to make future predictions. Markov chains are defined for both discrete and continuous stochastic processes. For our purpose, we focus only on the *Discrete time Markov chains.*

Definition 4.4 (Markov Chain). A sequence of S-valued random variables $X = \{X_i, i \in \mathbb{Z}\}$ is a *Markov Chain* if it satisfies the Markov property, that X_{n+1} depends only on X_n, that is,

$$P\left(X_{n+1} = x_{n+1} | X_0 = x_0, \ldots, X_n = x_n\right) = P\left(X_{n+1} = x_{n+1} | X_n = x_n\right),$$

where x_i is the state of the chain at time i.

We can characterize a Markov chain by three components namely: a state space $S = \{x_1, \ldots, x_k\}$, containing all the possible states X_n can take, an initial distribution p^0, quantifying the distribution of the starting point of the chain, and a transition probability or Markov kernel P, which details the probabilities of transitioning from a state x_i to x_j. Therefore, the transition probability P establishes the evolution of the Markov chain. For the rest of this section, we assume that the Markov chain is *homogeneous*, meaning that the transition probabilities are the same for all time period.

Thus, if we denote p_{ij} as the transition probability from state x_i to state x_j in one time step, then we can write p_{ij} as

$$p_{ij} = P\left(X_{n+1} = x_j | X_n = x_i\right).$$

When we write the above probabilities for all the different choices $1 \leq i, j \leq k$, the resulting transition matrix is given by

$$P = [p_{ij}], \quad 1 \leq i, j \leq k.$$

If we are interested in the probability of moving from a state x_i to state x_j in m-time steps, then, we denote this transition probability by

$$p_{ij}^{(m)} = P\left(X_{n+m} = x_j | X_n = x_i\right),$$

with the corresponding m-time step transition probability matrix

$$P_m = \left[p_{ij}^{(m)}\right] = P^m.$$

The initial density p^0 is given by

$$p^0 = \left[p_1^0, \ldots, p_k^0\right],$$

where $p_i^0 = P\left(X_0 = x_i\right), \quad i = 1, \ldots, k$. In addition, the entries of p^0 and P are non-negative since they contain probabilities. Thus, the elements of p^0 and rows of P must sum to unity.

Example 4.2. On a given day, Deeyah is either happy (H) or sad (S). Assume if she is happy today, the probability that she will be happy tomorrow is 0.7, whereas if she is sad today the chance of her being sad tomorrow is 60%. Thus, the probability of her being happy or sad on any particular day depends solely on whether she is happy or sad the previous day.

We can model her emotional state (Happy or Sad) as a two-state Markov chain where the state space is

$$S = \{\text{Happy, Sad}\},$$

with resulting transition probability matrix

$$P = \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}.$$

Furthermore,

$$p^0 = \begin{bmatrix} p_h^0 & p_s^0 \end{bmatrix}, \quad p_h^0 + p_s^0 = 1$$

is the percentage of time Deeyah is happy on any given day. To illustrate, assume

$$p^0 = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix}.$$

The probability that she will be happy tomorrow is given by

$$p^1 = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}$$

$$= \begin{bmatrix} 0.64 & 0.32 \end{bmatrix}.$$

Table 4.1: Distribution of π in Example 4.2

n	P^n	n	P^n	n	P^n
0	$\begin{bmatrix} 0.8000 & 0.2000 \end{bmatrix}$	4	$\begin{bmatrix} 0.5720 & 0.4267 \end{bmatrix}$	8	$\begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$
1	$\begin{bmatrix} 0.6400 & 0.3600 \end{bmatrix}$	5	$\begin{bmatrix} 0.5720 & 0.4280 \end{bmatrix}$	9	$\begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$
2	$\begin{bmatrix} 0.5920 & 0.4080 \end{bmatrix}$	6	$\begin{bmatrix} 0.5716 & 0.4284 \end{bmatrix}$	10	$\begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}$
3	$\begin{bmatrix} 0.5776 & 0.4224 \end{bmatrix}$	7	$\begin{bmatrix} 0.5714 & 0.4285 \end{bmatrix}$		

Thus, the distribution after n days is

$$p^n = \begin{bmatrix} 0.8 & 0.2 \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix}^n.$$

The distribution for $n = 0, 1, 2, \ldots, 10$ have been compiled in Table 4.1. The table shows that the distribution, π, is converging to a stationary value. The limiting distribution π can be solved by using

$$\pi = \pi P, \quad \sum \pi_i = 1$$

$$\Rightarrow \begin{bmatrix} \pi_{\text{happy}} & \pi_{\text{sad}} \end{bmatrix} \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} \pi_{\text{happy}} & \pi_{\text{sad}} \end{bmatrix}, \quad \pi_{\text{happy}} + \pi_{\text{sad}} = 1,$$

to obtain

$$\pi = \begin{bmatrix} 0.5714 & 0.4286 \end{bmatrix}.$$

We remark that in general we cannot always solve explicitly, for the stationary values of a Markov chain. Consider a Markov chain that converges in distribution to π, that is,

$$\lim_{n \to \infty} P^n = \pi.$$

Given that this limit exists, we must have that

$$\pi = \lim_{n \to \infty} p^0 P^n = \lim_{n \to \infty} p^0 P^{n+1} = \left(\lim_{n \to \infty} p^0 P^n \right) P = \pi P.$$

Definition 4.5 (Stationary Distribution). A Markov chain X_n with transition probability P and distribution π, is said to be stationary if it satisfies

$$\pi = \pi P, \quad \sum \pi_i = 1.$$

Remark 4.4. Every finite Markov chain has at least one stationary distribution. However, this distribution might not be unique and may not be equal to $\lim_{n\to\infty} P^n$.

Remark 4.5. In order to construct a Markov chain that has stationary distribution, we need to establish that the Markov chains are irreducible and aperiodic. We do not cover these concepts in details here. Interested readers should check, for instance [62, 78]. However, it is often difficult to solve for the stationarity of the chain using Definition 4.5, instead, we present the *detailed balance equation*, an alternative approach of establishing the stationarity of Markov chains.

Definition 4.6 (Detailed Balance Condition). A chain with transition matrix $P = [P_{ij}]$ and distribution $\pi = [\pi_1, \ldots, \pi_k]$ is reversible if the detailed balance equation

$$\pi_i P_{ij} = \pi_j P_{ji}$$

is satisfied for all i, j.

Since $\sum_i \pi_i P_{ij} = \sum_i \pi_j P_{ji} = \pi_j \sum_i P_{ji} = \pi_j$, it follows that $\pi P = \pi$, so the reversibility implies stationarity. We also observe that if the chains are irreducible and aperiodic, they will uniquely converge to a specific stationary distribution. Using Definition 4.6, it is clear that $\pi_{k-1} P_{k-1,k} = \pi_k P_{k,k-1}$ is sufficient but not a necessary requirement for the MCMC chains to be stationary.

Theorem 4.13 ([75], Section 8.4). *The MCMC chain of the local volatility parameter σ_{LV} is stationary.*

Proof. Let the posterior density at the k-th step be $\pi_k = \pi(\sigma_{LV}^k | V)$ and $P_{k-1,k} = P\left(X_k = \sigma_{LV}^k | X_{k-1} = \sigma_{LV}^{k-1}\right)$ to be the transition probability from σ_{LV}^{k-1} to σ_{LV}^k. To show stationarity, by Definition 4.6, we need to prove that

$$\pi\left(\sigma_{LV}^{k-1}|V\right) P_{k-1,k} = \pi\left(\sigma_{LV}^k|V\right) P_{k,k-1}.$$

Since $P_{k-1,k} = P\left(\text{proposing } \sigma_{LV}^k\right) P\left(\text{accepting } \sigma_{LV}^k\right)$. It then follows from the definition of proposal density and acceptance probability α, that

$$
\begin{aligned}
P_{k-1,k} &= J\left(\sigma_{LV}^k|\sigma_{LV}^{k-1}\right) \alpha\left(\sigma_{LV}^k|\sigma_{LV}^{k-1}\right) \\
&= J\left(\sigma_{LV}^k|\sigma_{LV}^{k-1}\right) \min\left(1, \frac{\pi\left(\sigma_{LV}^k|V\right) J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^k\right)}{\pi\left(\sigma_{LV}^{k-1}|V\right) J\left(\sigma_{LV}^k|\sigma_{LV}^{k-1}\right)}\right).
\end{aligned}
\tag{4.22}
$$

Note that if the jumping density is symmetric, that is, $J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^{k}\right) = J\left(\sigma_{LV}^{k}|\sigma_{LV}^{k-1}\right)$, then $\alpha\left(\sigma_{LV}^{k}|\sigma_{LV}^{k-1}\right) = \min\left(1, \frac{\pi\left(\sigma_{LV}^{k}|V\right)}{\pi\left(\sigma_{LV}^{k-1}|V\right)}\right)$. To further simplify (4.22), we need the following relation whose proof can be found in Appendix A. For $x, v \in \mathbb{R}$,

$$v \min(1, x/v) = \min(x, v) = x \min(1, v/x). \tag{4.23}$$

Using (4.23), it follows that

$$
\begin{aligned}
&\pi\left(\sigma_{LV}^{k-1}|V\right) P_{k-1,k} \\
&= \pi\left(\sigma_{LV}^{k-1}|V\right) J\left(\sigma_{LV}^{k}|\sigma_{LV}^{k-1}\right) \min\left(1, \frac{\pi\left(\sigma_{LV}^{k}|V\right) J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^{k}\right)}{\pi\left(\sigma_{LV}^{k-1}|V\right) J\left(\sigma_{LV}^{k}|\sigma_{LV}^{k-1}\right)}\right) \\
&= \pi\left(\sigma_{LV}^{k}|V\right) J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^{k}\right) \min\left(1, \frac{\pi\left(\sigma_{LV}^{k-1}|V\right) J\left(\sigma_{LV}^{k}|\sigma_{LV}^{k-1}\right)}{\pi\left(\sigma_{LV}^{k}|V\right) J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^{k}\right)}\right) \\
&= \pi\left(\sigma_{LV}^{k}|V\right) P_{k,k-1}.
\end{aligned}
\tag{4.24}
$$

Hence, the detailed balance condition is satisfied for the Metropolis Hastings algorithm and the posterior density is the stationary distribution. $\square$

4.3.2 Markov Chain Monte Carlo (MCMC) Methods

In most parameter estimation problems, the posterior joint density of the parameters is defined in $\mathbb{R}^p$. The evaluation of this posterior distribution for moderate dimensionality p involves the use of quadrature techniques (especially sparse grid quadrature) [75]. For large dimensionality p, Monte Carlo integration techniques are often used. The challenge with using Monte Carlo approach is that the support of the density is often part of the information that we are seeking. Therefore, instead of specifying some parameter values and evaluating the posterior density at those values, we instead seek a better alternative in exploring the geometry of the density by specifying parameter values using the properties of the density.

So, we construct a Markov chain whose stationary distribution coincides with the posterior density of the parameter. The idea behind the MCMC techniques involves simulating the realizations of the Markov chains from the posterior density. These chains give the parameter values that are sampled from the parameter density given the observed data samples.

4.3.3   Metropolis Hasting Algorithm

We now turn to the strategy by which we simulate the Markov chains. Recall that the next state of a Markov chain only depends on the present state. Given this information, we construct the Markov chain as follows:

Strategy 4.1. Consider the parameter $\sigma_{LV}^{k-1} \in \mathbb{R}^p$ has been specified.

1. Assign the current chain realization to σ_{LV}^{k-1}, that is, $X_{k-1} = \sigma_{LV}^{k-1}$.

2. Propose a new value $\sigma_{LV}^* \sim J\left(\sigma_{LV}^*|\sigma_{LV}^{k-1}\right)$ based on the jumping or proposal distribution J. So, we specify the next value σ_{LV}^* of the chain based on the previous value σ_{LV}^{k-1}, but this should not be understood as a conditional density.

3. With probability $\alpha\left(\sigma_{LV}^*|\sigma_{LV}^{k-1}\right)$ computed by using the posterior distribution, accept σ_{LV}^{k-1} and assign it to X_k; i.e., $X_k = \sigma_{LV}^*$. Otherwise, take $X_k = \sigma_{LV}^{k-1}$. Also, $\alpha\left(\sigma_{LV}^*|\sigma_{LV}^{k-1}\right)$ should not be interpreted as a conditional probability but rather as a probability of accepting σ_{LV}^* generated from the previous value σ_{LV}^{k-1}.

4. Establish that the stationary distribution for the chain is the posterior density.

From the strategy above, we construct a *Metropolis Hasting Algorithm*. Another variation of this algorithm which utilizes this strategy is *Random Walk Metropolis*, see [75] for more details. In the context of the work presented in Section 4.4, we consider the case where the proposal or jumping density is taken to be symmetric such that $J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^k\right) = J\left(\sigma_{LV}^k|\sigma_{LV}^{k-1}\right)$. In this respect, we consider two possibilities for the jumping distribution:

$$J\left(\sigma_{LV}^*|\sigma_{LV}^{k-1}\right) = \mathcal{N}\left(\sigma_{LV}^{k-1}, V\right), \text{ and } J\left(\sigma_{LV}^*|\sigma_{LV}^{k-1}\right) = \mathcal{N}\left(\sigma_{LV}^{k-1}, D\right). \quad (4.25)$$

Here, V and D represent the variance-covariance and variance matrices for $\sigma_{LV} \in \mathbb{R}^p$ respectively. By the symmetric property of the proposal density, it follows

that

$$J\left(\sigma_{LV}^{*}|\sigma_{LV}^{k-1}\right) = \frac{1}{(2\pi)^p|V|}e^{-\frac{1}{2}\left[\left(\sigma_{LV}^{*}-\sigma_{LV}^{k-1}\right)V^{-1}\left(\sigma_{LV}^{*}-\sigma_{LV}^{k-1}\right)\right]}$$

$$= \frac{1}{(2\pi)^p|V|}e^{-\frac{1}{2}\left[\left(\sigma_{LV}^{k-1}-\sigma_{LV}^{*}\right)V^{-1}\left(\sigma_{LV}^{k-1}-\sigma_{LV}^{*}\right)\right]} \qquad (4.26)$$

$$= J\left(\sigma_{LV}^{k-1}|\sigma_{LV}^{*}\right).$$

Replacing V with D in (4.26) gives the same result. In addition, one can also use *t-distributed* proposal density for smaller sampled observations. The t-distribution is also a symmetric distribution. To see this, assume $T \sim t_r$ is t-distributed random variable with r degrees of freedom, then,

$$T = \frac{Z}{\sqrt{\chi/n}},$$

where Z and χ are independent standard normal and chi-squared random variables respectively. If T is symmetric, then $P(T > c) = P(T < -c)$. Thus,

$$\begin{aligned}
P(T > c) &= P\left(\frac{Z}{\sqrt{\chi/n}} > c\right) \\
&= P\left(Z > c\sqrt{\chi/n}\right) \\
&= P\left(Z > c\sqrt{\chi/n}|X\right)P(X) \\
&= P\left(Z < -c\sqrt{\chi/n}|X\right)P(X) \qquad (4.27) \\
&= P\left(Z < -c\sqrt{\chi/n}\right) \\
&= P\left(\frac{Z}{\sqrt{\chi/n}} < -c\right) \\
&= P(T < -c),
\end{aligned}$$

where in line three and five, we have used the property of conditional probability; in line four the symmetric property of normal distribution.

Next, we present the Metropolis Hasting Algorithm and discuss the motivation behind the use of symmetric proposal density in sampling the local volatility parameter.

We remark some key observations of Algorithm 4.1.

Algorithm 4.1 Metropolis Hasting Algorithm in Bayesian framework

1. Choose initial parameter $\vec{\sigma}^0$ such that $\pi\left(\vec{\sigma}^0|V\right) > 0$; $\dim(\vec{\sigma}^0) = p$.

2. For $k = 1, \ldots, M$

   (a) for $z \sim \mathcal{N}(0, \mathcal{I}_p)$, construct the candidate

   $$\vec{\sigma}^* = \vec{\sigma}^{k-1} + Rz,$$

   where $RR^T = V$ (Cholesky decomposition) for the covariance matrix V of the parameter $\vec{\sigma}$. Sometimes V is replaced with the diagonal matrix D whose elements contains the variances of each parameter in $\vec{\sigma}$.

   (b) $\vec{\sigma}^* \sim \mathcal{N}(\vec{\sigma}^{k-1}, V)$.

   (c) Compute the probability $r(\vec{\sigma}^*|\vec{\sigma}^{k-1})$ such that

   $$r(\vec{\sigma}^*|\vec{\sigma}^{k-1}) = \frac{\pi(\vec{\sigma}^*|V)}{\pi(\vec{\sigma}^{k-1}|V)} = \frac{\pi(V|\vec{\sigma}^*)\pi_0(\vec{\sigma}^*)}{\pi(V|\vec{\sigma}^{k-1})\pi_0(\vec{\sigma}^{k-1})};$$

      i. set $\vec{\sigma}^k = \vec{\sigma}^*$ with probability $\alpha = \min(1, r)$;

      ii. Otherwise $\vec{\sigma}^k = \vec{\sigma}^{k-1}$.

Remark 4.6.   1. First we note that, by taking the ratio of the posterior density, we eliminate the normalization constants in the acceptance probability of Algorithm 4.1, thereby reducing the computation.

2. Using the likelihood function of the normal distributed errors and similarly for t-distributed errors in Section 4.4, we eliminate the potential numerical $\frac{0}{0}$ evaluation.

3. A candidate value σ_{LV}^* that yields $\pi(V|\vec{\sigma}^*) > \pi(V|\vec{\sigma}^{k-1})$ leads to a smaller sums of squared errors allowing such candidate to be accepted with probability one.

4. If σ_{LV}^* is selected such that the sums of squared errors is increased, i.e., $\pi(V|\vec{\sigma}^*) < \pi(V|\vec{\sigma}^{k-1})$, then we accept this candidate with probability α.

4.3.4   Convergence Analysis of MCMC

Theorem 4.13 ensures that if the MCMC chains are run *sufficiently long enough,* they will produce samples from the posterior density. However, there is still question of determining how long to run the chain for this to happen. In other words, how long is sufficient to run the chain so that it converges and adequately samples from the posterior density. In general, this is quite difficult to answer as analytic convergence and stopping criteria are lacking for this procedure. In fact, some authors have noted that the convergence or burn-in of MCMC algorithms can be falsified but in general, not completely verified, see [15] for instance.

We now turn to answer the second question. Below, we list some numerical methods that can be used to test for the convergence of the MCMC chains. These are taken from several monographs such as [15, 45, 75].

1. One can visually observe that the marginal chains of the MCMC for each parameter is stationary after running for sufficiently long period of time. However, this raises a question in the case where the MCMC chain initially samples from a local minimum before determining another step with a lower residual. In fact, there is no way to guarantee this is the global

71

minimum and if the chain will transition again if a new lower minimum is found.

2. For moderate number of parameters, the MCMC chains can be compared with sample estimates from using quadrature techniques. We might add that using quadrature rule is not suitable for the local volatility model due to the higher dimensionality of the calibrated parameter.

3. From a statistical point of view, the acceptance ratio, α, can be used to determine if the chain is adequately sampling from the posterior density. The optimal acceptance ratio ranges from 0.1 to 0.5.

4. The use of autocorrelation between different component in the chain that are k-iterations apart can be of importance in determining covergence. The autocorrelation is given by

$$R(k) = \frac{\sum_{i=1}^{M-k}(\sigma_{LV}^i - \bar{\sigma}_{LV})(\sigma_{LV}^{i+k} - \bar{\sigma}_{LV})}{\sum_{i=1}^{M}(\sigma_{LV}^i - \bar{\sigma}_{LV})^2} = \frac{\text{Cov}(\sigma_{LV}^i, \sigma_{LV}^{i+k})}{\text{Var}(\sigma_{LV}^i)}.$$

Since adjacent components are likely correlated due to markov property, autocorrelation is used to determine if the chain is producing i.i.d. samples from the posterior density. In fact, low autocorrelation is often indicative of fast convergence [16].

5. Convergence can also be monitored to check if the MCMC has run long enough by checking the *Potential Scale Reduction Factor* (PSRF) of the estimates of call option prices [40, 45].

To calculate PSRF, we consider the following procedure. Suppose we have run the MCMC for m chains with lengths n (i.e length of each chain) after the burn in period. For each calibration, denote V_{ij} the option price of the ith parameter of the jth simulated chain, where $i = 1, \ldots, n$ and $j = 1, \ldots, m$. Furthermore, we define the between variances, B, and within-sequence variances, W, as follows:

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\bar{V}_{.j} - \bar{V}_{..})^2 \quad \text{and} \quad W = \frac{1}{m}\sum_{j=1}^{m}s_j^2,$$

where

$$\bar{V}_{.j} = \frac{1}{n}\sum_{i=1}^{n}V_{ij}, \quad \bar{V}_{..} = \frac{1}{m}\sum_{i=1}^{m}\bar{V}_{.j}, \quad s_j^2 = \frac{1}{n-1}\sum_{j=1}^{n}(V_{ij} - \bar{V}_{.j})^2.$$

72

With the above given definitions, we compute PSRF as follows:

$$\text{PSRF}(V) = \sqrt{1 - \frac{1}{n}\left(1 - \frac{B}{W}\right)}. \qquad (4.28)$$

The convergence of the chain is observed by assessing the scale factor by which the current distribution for V might be reduced as the length of the chain gets large. Notice that PSRF $\rightarrow 1$ as $n \rightarrow \infty$. If PSRF is high, then it means continuing simulation would improve convergence. In otherwords, it signifies that convergence can not be yet ascertained.

The PSRF estimate works because B which represents the between sequence variances usually overestimates posterior variance, whereas W usually underestimates the posterior variances. Moreover, as we run the chain for longer period, B gets closer to W, hence, the ratio B/W gets closer to 1. So, the PSRF estimate for V given by (4.28) goes to 1 too. See [41, 40] for further references on PSRF values. It is noted that a good estimate range for PSRF is between 0.9 to 1.1.

4.4   Bayesian Modeling

In this section, we explore the calibration of volatility parameter via four different statistical models, namely: additive normally distributed error, additive student-t distributed error, multiplicative normally distributed error, and multiplicative student-t distributed error models. In the context of Bayesian learning, the previous literature on extracting local volatility surface are very much minimal. More recently, [45] worked on Bayesian method of sampling the volatility parameter from its posterior distribution. However, their implementation differs from the work presented here in two ways. First, we compute local volatility from four different statistical models without restricting to only *normally* distributed additive error model as in [45]. Second, our choice of the prior density for the local volatility parameter σ_{LV} which is incorporated into the Bayesian framework of the four statistical error models under investigation, differs from theirs, and hence, the posterior distribution of σ_{LV} differs. Furthermore, we reformats the calibration problem into Bayesian framework by seeking to obtain the posterior distribution for the local volatility parameter via these models.

Thus, in this section, we develop mathematical tools necessary for the calibration of the statistical models in estimating the unknown parameter σ_{LV} that is consistent with the dynamics of the asset prices by treating it as a random variable for each exercise price and maturity.

To begin with, we give a general set up for the additive error model. Consider

$$\Upsilon_i = f(t_i, q_0) + \epsilon_i, \qquad i = 1, 2, 3, \ldots, n, \tag{4.29}$$

where Υ_i are random variables whose realizations v_i are the sets of measurement; $f(t_i, q_0)$ is the model response parametrized by q_0 at the corresponding times t_i. There are various ways one can quantify q_0 such that the model response and the set of observations are as close as possible within a given tolerance level. Here, we quantify such q_0 by obtaining its distribution. The so-called "posterior" density for the parameter involves the probability density under which the v_i are the most likely to be observed. We denote it by $\pi(q|v)$. In the context of parameters that are to be quantified based on observations, the posterior density can be characterized completely by applying the *Bayes' rule*:

$$\pi(q|v) = \frac{\pi(v|q)\pi_0(q)}{\pi_\Upsilon(v)}, \tag{4.30}$$

where $\pi_0(q)$ represents the *prior* density and $\pi(v|q)$ is the likelihood function that describes how likely the observations are, given the parameter q. The denominator $\pi_\Upsilon(v)$ in (4.30) is a normalizing constant given by

$$\pi_\Upsilon(v) = \int_{\mathbb{R}^p} \pi(v|q)\pi_0(q)dq,$$

which in practice is not calculated.

Now, consider a financial derivative (specifically a European option) whose value depends on the prices of the underlying asset over time. The dynamics of the asset prices are assumed to be dependent on an unknown parameter σ_{LV} from the prescribed local volatility model. Suppose that, at time $t \in \Upsilon_n([0, T]) = \{0 \le t_1, t_2, \ldots, t_n \le T\}$ we have a set of observed European option prices across different strikes K_i and maturities T_i, where $i \in I_t$ is an index set. Hence, the objective is to quantify the volatility parameter that is consistent with the observed option prices $V_t = \{V_t^{(i)}, i \in I_t, t \in \Upsilon_n\}$ under local volatility modeling.

To simplify the framework, we use an average of the quoted bid-ask spread

$$V_t^{(i)} = \frac{1}{2} \left(V_t^{(i)\text{bid}} + V_t^{(i)\text{ask}} \right), \tag{4.31}$$

for each $i \in I_t$.

In order to obtain a complete representation for the posterior density $\pi(q|v)$, one needs to completely express the likelihood function $\pi(v|q)$ and the prior distribution $\pi_0(q)$. We incorporate the set of prior beliefs in Section 2.3.2 into the prior density of the parameter.

To set up the likelihood function, we model the noise from a normal distribution with mean zero and fixed but unknown variance δ_0^2. For a set of option prices $V_0^{(i)}$ for every $i \in I_0 = \{1, \ldots, n\}$ as in (4.31) observed today at time $t = 0$, we define $\delta_i = \frac{1}{2} \left| V_0^{(i)\text{bid}} - V_0^{(i)\text{ask}} \right|$ to be the basis point bid-ask spread for the ith option at time zero. Here δ_i can be interpreted as the deviation from the mean zero. Having defined the sum of squares

$$SS(\sigma_{LV}) = \sum_{i=1}^{n} \left| V_0^{(i)} - f_i(\sigma_{LV}) \right|^2$$

for the quoted values $V_0^{(i)}$ and the theoretical values $f_i(\sigma_{LV})$ from the local volatility model, we assign positive likelihood under σ_{LV} only if

$$\left| V_0^{(i)} - f_i(\sigma_{LV}) \right| \leq \delta_i, \quad \text{for all} \quad i = 1, \ldots, n. \tag{4.32}$$

In order to propose a more efficient and robust numerical procedure, we choose δ such that all the option prices at time zero fall within this tolerance level. Specifically,

$$SS(\sigma_{LV}) \leq \delta^2 = \sum_{i=1}^{n} \delta_i^2. \tag{4.33}$$

In the remaining part of this section, we derive a mathematical formula for computing local volatilities based on various statistical models. These models include: *normally* distributed additive error, *student-t* distributed additive error, *normally* distributed multiplicative error, and *student-t* distributed multiplicative error models.

4.4.1 Normally distributed additive error model

First, we explore the statistical model with normally distributed additive errors. The model is given as:

$$V_i = f_i(\sigma_{LV}) + \epsilon_i, \qquad i = 1, 2, 3, \ldots, n, \tag{4.34}$$

with $\epsilon_i \sim \mathcal{N}(0, \delta^2)$ for $i = 1, \ldots, n$ and $\delta^2 = \sum_{i=1}^n \delta_i^2$. Here, the δ_i are chosen as $\delta_i = \frac{1}{2} \left| V_0^{(i)\text{bid}} - V_0^{(i)\text{ask}} \right|$.

Under the i.i.d. assumptions for the error, the likelihood function $\pi(V|\sigma_{LV}, \delta^2)$ is given by

$$\pi(V|\sigma_{LV}, \delta^2) \propto \mathbb{1}_{\{SS(\sigma_{LV}) \leq \delta^2\}} \exp\left\{ -\frac{1}{2\delta^2} SS(\sigma_{LV}) \right\}, \tag{4.35}$$

where $\mathbb{1}$ represents the indicator function under which the likelihood function is positive and $\text{SS}(\sigma_{LV}) = \sum_{i=1}^n \left| V_0^{(i)} - f_i(\sigma_{LV}) \right|^2$. Furthermore, the likelihood function is constructed via the densities of the i.i.d. errors. For more details on the construction of the likelihood function, we refer the reader to [76, Section 8.1]. Using the prior density constructed in [45] we have

$$\pi_0(\sigma_{LV}) \propto \exp\left\{ -\frac{1}{2} \|\log(\sigma_{LV}) - \log(\sigma_{imp})\|_\kappa^2 \right\}, \tag{4.36}$$

where the norm $\|\cdot\|_\kappa$ is defined as

$$\|u\|_\kappa^2 = (1 - \kappa) \|u\|_2^2 + \kappa \||\nabla u|\|_2^2.$$

Here, $\nabla = \left(\frac{\partial}{\partial T}, \frac{\partial}{\partial K} \right)$ is the gradient operator with $|\nabla u| = \left(\frac{\partial u}{\partial T} \right)^2 + \left(\frac{\partial u}{\partial K} \right)^2$ and $\|\cdot\|_2$ is the standard L_2-norm of the square integrable functions; $\kappa \in (0, 1)$ is a pre-specified constant.

Combining the likelihood function and prior density gives the posterior density

$$\pi(\sigma_{LV}|V) \propto \mathbb{1}_{\{SS(\sigma_{LV}) \leq \delta^2\}}$$
$$\times \exp\left\{ -\frac{1}{2\delta^2} \left[SS(\sigma_{LV}) + \tilde{\lambda} \|\log(\sigma_{LV}) - \log(\sigma_{imp})\|_\kappa^2 \right] \right\}, \tag{4.37}$$

where $\tilde{\lambda} = \delta^2$ is the penalty parameter.

In the next section, we propose the use of student-t distribution for the errors, especially for the case where we have few observations ($n < 30$) in the market. We also recommend the use of student-t for the error distribution when working with sampled covariance matrix rather than known population covariance [68].

4.4.2 Student-t distributed additive error model

Second, we consider the statistical model:

$$V_i = f_i(\sigma_{LV}) + \tau_i, \text{ for } i = 1, 2, 3, \ldots, n, \tag{4.38}$$

where the τ_i are i.i.d. t-distributed random variables with $n - p$ degrees of freedom. Note that p is the size of the parameter q while n is the number of observed option prices. Note that the degrees of freedom for this error modeling depends on the number p of parameters σ_{LV} chosen on the surface to estimate. The value of p is estimated based on the trade-off between model accuracy and computational time. This could also be estimated via Bayesian framework by treating it as an unknown and assuming a reasonable prior. However, a suitable range of values for p satisfy the conditions above and provides a smooth surface. Here, we assume for simplicity that $p = 89$ as a pre-estimated value. This yields a fixed number of degrees of freedom $(n - p)$ for the random variable τ. Within the model, the density of τ with $k = n - p$ is given by

$$\pi_k(\tau) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi k}} \left(1 + \frac{1}{k}(V - f(\sigma_{LV}))^2\right)^{-\frac{(k+1)}{2}}.$$

Therefore, the likelihood function can be constructed by using the i.i.d. property as well as the densities of the errors:

$$
\begin{aligned}
\pi(V|\sigma_{LV}, k) &= \left(\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi k}}\right)^n \prod_{i=1}^n \left(1 + \frac{1}{k}(V_i - f_i(\sigma_{LV}))^2\right)^{-\frac{(k+1)}{2}} \\
&\propto \prod_{i=1}^n \left(1 + \frac{1}{k}(V_i - f(\sigma_{LV}))^2\right)^{-\frac{(k+1)}{2}} \\
&\propto \exp\left\{\sum_{i=1}^n \log\left(1 + \frac{1}{k}(V_i - f_i(\sigma_{LV}))^2\right)^{-\frac{(k+1)}{2}}\right\} \\
&\propto \exp\left\{-\frac{(k+1)}{2}\sum_{i=1}^n \log\left(1 + \frac{1}{k}(V_i - f_i(\sigma_{LV}))^2\right)\right\}.
\end{aligned}
\tag{4.39}
$$

Combining the prior distribution and the likelihood gives the posterior distribution as

$$\pi(\sigma_{LV}|V, k) \propto \exp\left\{-\frac{(k+1)}{2}\left[\sum_{i=1}^n \log\left(1 + \frac{1}{k}(V_i - f_i(q))^2\right) + \tilde{\lambda}\left\|\log\left(\frac{\sigma_{LV}}{\sigma_{imp}}\right)\right\|_\kappa^2\right]\right\},$$
$$\tag{4.40}$$

where $\tilde{\lambda} = \frac{1}{k+1}$.

In the next two sections, we pass to multiplicative statistical models. For a structural model, the distribution of the residuals from additive error models are highly dependent on the magnitude of the observations even though the model provides accurate fit to the data. So, for this case, a multiplicative error distribution is more appropriate as the variance of the observation will depend on the magnitude of the model response.

The local volatility model exhibits the properties of a structural model, hence, the motivation for its application here.

4.4.3 Normally distributed multiplicative error model

Third, we impose the following statistical model for estimating local volatility parameter:

$$V_i = f_i(\sigma_{LV})\left(1 + \epsilon_i\right), \qquad i = 1, 2, 3, \ldots, n, \qquad (4.41)$$

where the model errors are i.i.d. and

$$\epsilon_i \sim \mathcal{N}(0, \delta^2), \quad i = 1, 2, 3, \ldots, n.$$

Therefore, it follows from (4.41) that

$$V_i \sim \mathcal{N}\left(f_i(\sigma_{LV}), f_i(\sigma_{LV})^2\delta^2\right).$$

By using the i.i.d. property and densities of the errors, the likelihood function in this setting is given by

$$\pi(V|\sigma_{LV}, \delta^2) \propto \mathbb{1}_{\{SS(\sigma_{LV}) \leq f(\sigma_{LV})^2\delta^2\}} \exp\left\{-\frac{1}{2f(\sigma_{LV})^2\delta^2}SS(\sigma_{LV})\right\}. \qquad (4.42)$$

With this likelihood function and the initial prior in (4.36), we obtain the posterior distribution as

$$\pi(\sigma_{LV}|V) \propto \mathbb{1}_{\{SS(\sigma_{LV}) \leq f(\sigma_{LV})^2\delta^2\}}$$

$$\times \ \exp\left\{-\frac{1}{2f(\sigma_{LV})^2\delta^2}\left[SS(\sigma_{LV}) + \tilde{\lambda}\left\|\log\left(\frac{\sigma_{LV}}{\sigma_{imp}}\right)\right\|_\kappa^2\right]\right\}, \quad (4.43)$$

where $\tilde{\lambda} = f(\sigma_{LV})^2\delta^2$ is again a penalty parameter.

4.4.4 Student-t distributed multiplicative error model

Finally, we use the following statistical model:

$$V_i = f_i(\sigma_{LV})\,(1 + \tau_i), \qquad i = 1, 2, 3, \ldots, n, \tag{4.44}$$

where the τ_i are i.i.d. t-distributed random variables with $n - p$ degrees of freedom. The reason behind the choice of the degrees of freedom here is similar to the discussion in student-t distributed additive error model. Hence, the density of τ with $k = n - p$ is given as

$$\pi_k(\tau) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi k}f(\sigma_{LV})}\left(1 + \frac{1}{k}\left(\frac{V - f(\sigma_{LV})}{f(\sigma_{LV})}\right)^2\right)^{-\frac{(k+1)}{2}}.$$

The likelihood function is then obtained as

$$\pi(V|\sigma_{LV}, k) = \left(\frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+1}{2}\right)\sqrt{\pi k}}\right)^n \prod_{i=1}^{n}\frac{1}{f_i(\sigma_{LV})}\left(1 + \frac{1}{k}\left(\frac{V_i - f_i(\sigma_{LV})}{f_i(\sigma_{LV})}\right)^2\right)^{-\frac{(k+1)}{2}}$$

$$\propto \prod_{i=1}^{n}\frac{1}{f_i(\sigma_{LV})}\left(1 + \frac{1}{k}\left(\frac{V_i - f_i(\sigma_{LV})}{f_i(\sigma_{LV})}\right)^2\right)^{-\frac{(k+1)}{2}}$$

$$\propto \exp\left\{-\sum_{i=1}^{n}\log(f_i(\sigma_{LV}))\right.$$

$$\left. + \sum_{i=1}^{n}\log\left(1 + \frac{1}{k}\left(\frac{V_i - f_i(\sigma_{LV})}{f_i(\sigma_{LV})}\right)^2\right)^{-\frac{(k+1)}{2}}\right\}$$

$$\propto \exp\left\{-\sum_{i=1}^{n}\log(f_i(\sigma_{LV}))\right.$$

$$\left. -\frac{(k+1)}{2}\sum_{i=1}^{n}\log\left(1 + \frac{1}{k}\left(\frac{V_i - f_i(\sigma_{LV})}{f_i(\sigma_{LV})}\right)^2\right)\right\}.$$

$$\tag{4.45}$$

Combining (4.45) with the initial prior in (4.36) yields the posterior distribution

$$\pi(\sigma_{LV}|V, k) \propto \exp\left\{-\sum_{i=1}^{n}\log(f_i(\sigma_{LV})) - \frac{1}{2}\left\|\log\left(\frac{\sigma_{LV}}{\sigma_{imp}}\right)\right\|_{\kappa}^2\right.$$

$$\left. -\frac{(k+1)}{2}\sum_{i=1}^{n}\log\left(1 + \frac{1}{k}\left(\frac{V_i - f_i(\sigma_{LV})}{f_i(\sigma_{LV})}\right)^2\right)\right\}. \tag{4.46}$$

Remark 4.7. In all the four statistical error models we have discussed, one may choose a *flat prior*

$$\pi_0(\sigma_{LV}) = \mathbb{1}_{\sigma_{LV}\geq 0} \tag{4.47}$$

and compute the posteriors respectively.

4.5   Numerical Example

Having built the mathematical tools necessary to obtain the local volatility surfaces, we now analyze and compare the results from each model. The data used to test the algorithms is taken from [2]: it is a collection of bid-ask spreads of European call options on S&P 500 stock index. The spread consists of 163 observations across 18 strikes and 12 maturities. The raw data is refined by averaging the bid and ask implied volatilities for each strike K and maturity T.

4.5.1   Results and Analysis

Now, we introduce the MCMC algorithm for sampling the local volatility parameter from the posterior distribution. The generic procedure is given in Algorithm 4.1. For the MCMC algorithm, we run a chain of 10000 surfaces including a burn-in period of 500. Estimates of σ at the discretized points are sampled at each iteration of the algorithm. These points are chosen across all strikes and maturities from the observed 163 market option prices. The number of discretized points for the volatility parameter is assumed to be less than the number of observed option prices used (for more details, see [76, Section 7.2]). Here, we choose 89 as the pre-estimated number of parameters. Consequently, an increase in the number of discretized points improves the stability of the method; however, it also slows down the procedure. Therefore, a trade-off between robustness and computational efficiency has to be taken into account.

In the following, we present the local volatility surfaces and relative norms of the MCMC chain. Figure 4.1 is constructed using Algorithm 4.1 in the Bayesian framework. The volatility surfaces obtained are modeled from the normally distributed additive error model, the student-t distributed additive error model, the normally distributed multiplicative error model, and the student-t distributed multiplicative error model. Since changes only appear at the beginning including the burn in period of the MCMC, we depict Figure 4.1 (right column) restricting

80

the x-axis to 500 iterations. In particular, after the 500th iteration, there are no significant changes in either the relative norms of the errors in the MCMC chain or the surfaces obtained at each iteration. However, we continued the MCMC until the 10000th iteration and depict the surfaces in Figure 4.1 (left column). The discretization of the parameter space $\sigma_{LV}(T, K)$ can be found in [45].

To start with, we discuss the properties of the volatility surfaces via the Bayesian method. Unlike the classical methods [17, 5], we do not encounter negative local volatility in the Bayesian framework due to a suitable choice of the prior density for the parameter σ. Also, we ensure smoothness in the local volatility surfaces obtained under the Bayesian method. This is done by incorporating the *smoothness* property (c) on page 14 into the prior density of the parameter σ. By construction, the values of the tolerance levels δ_i for the replication of model prices $f_i(\sigma_{LV})$ are directly proportional to the value of δ. So, as the δ_i get smaller, δ (that is, higher resolution of volatility surface) gets small; hence, the value of δ is dependent on the data sample. This is crucial as the smoothness parameter for the normally distributed error models, $\tilde{\lambda}$ depends on the value of δ. Hence, as p (the number of chosen points in the parameter space) gets large, the value of δ increases. This makes sense since the more points chosen in the parameter space means a higher resolution for the resultant volatility surface (as discretization is a form of regularization); hence, a bigger value of δ becomes a good choice. Similarly, for the student-t distributed error models, the smoothness parameter $\hat{\lambda} = \frac{1}{n-p+1}$ also depends on the value of p. Therefore, as the number of points chosen on the parameter grid increases, the value of the smoothness parameter increases. Converse is also true for small values of p. Furthermore, we can easily see the volatility frown (sneer) produced under the Bayesian framework in Figure 4.1 at low maturities and strike prices. For normally distributed additive error model, the volatility skewness can be seen clearly in Figure 4.2 for different strike and maturity values.

Next, we compare and contrast the surfaces obtained via the four models discussed in Section 4.4. Here, we first discuss the pros and cons of the additive error models. The volatility surface obtained from normally distributed additive error model is similar to that of student-t distributed additive error model.

81
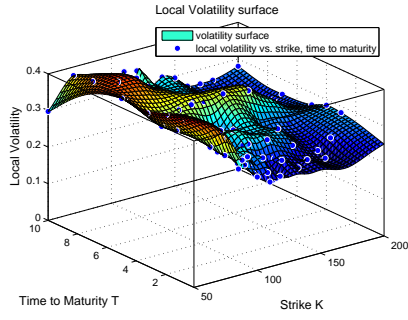
(a) normally distributed additive error
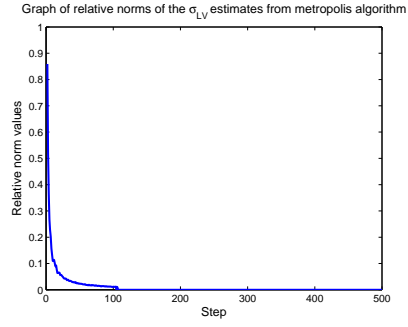
(b) relative norms of MCMC chain
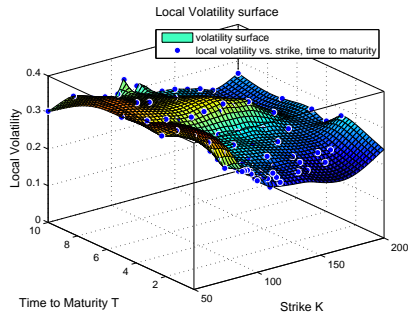
(c) student-t distributed additive error

(d) relative norms of MCMC chain

(e) normally distributed multiplicative error

(f) relative norms of MCMC chain

(g) student-t distributed multiplicative error

(h) relative norms of MCMC chain

Figure 4.1: Local volatility surface via Bayesian framework; the dots on the surfaces represent the estimates at the chosen points for the local volatility parameter:

82

(a) Local volatility skew for $T = 0.08$ and $T = 6$.
(b) Local volatility skew for $K = 50$ and $K = 100$.

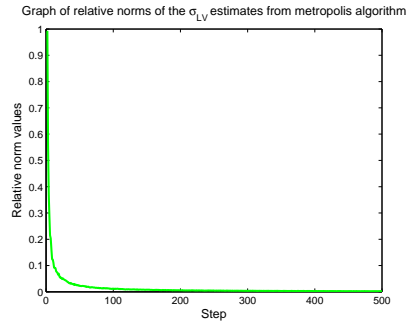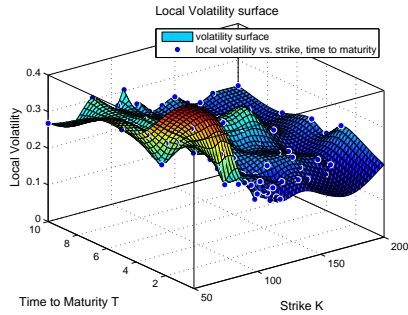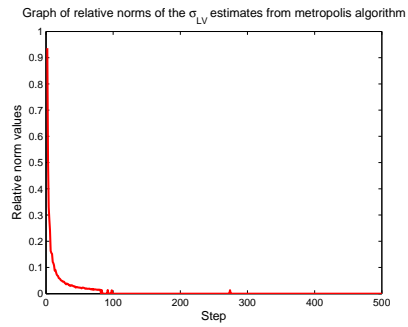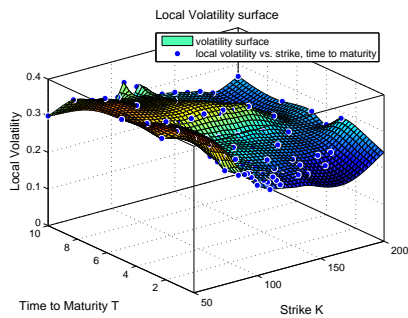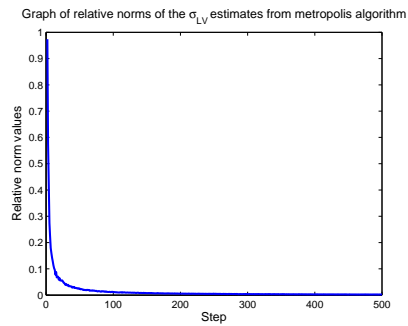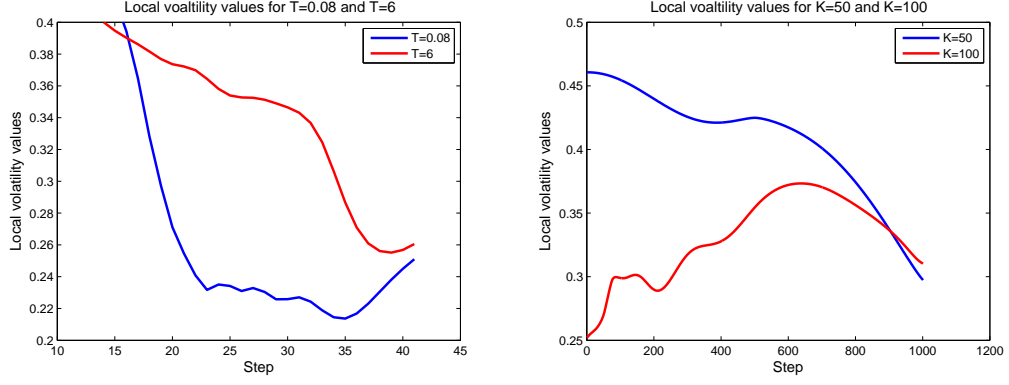Figure 4.2: Local volatility smile and frowns for normally distributed additive error model

However, in the context of student-t distributed additive error model, the use of sample covariance matrix in place of unknown population covariance matrix of the parameter σ is statistically accurate compared to the normally distributed additive error model. In addition, the relative norms of the MCMC chains converge to zero quickly in both additive error models. Also, student-t distributed additive error model has higher sums of squared error of 1642.3 compared to 1585.2 of normally distributed additive error model.

On the other hand, we also analyze the surfaces obtained via the multiplicative error models. The normally distributed multiplicative error model has the fastest convergence of relative norms of its MCMC chain among all the four models. However, normally distributed multiplicative error model also has the worst sum of squared error of 2242.9 among the four models, while student-t distributed multiplicative error model has 1633.7 sum of squared errors.

In the case of Bayesian approach, it may not be possible to distinguish between the volatility surfaces in order to propose to a practitioner. However, within a prescribed confidence interval, a distribution of volatility surfaces can be selected. This confidence interval can be constructed using the estimated covariance matrix of σ_{LV}. Furthermore, given the analysis of the surfaces in this work based on the data used, a financial analyst can select one of the sur-

faces in Figure 4.1 depending on his/her preferences. To be more specific, if the preference of an analyst is to choose a surface with the least sums of squared errors, then the volatility surface from normally distributed *additive* error model should be preferred. However, if the rate of convergence of the MCMC chains is an important factor, then the volatility surface from normally distributed *multiplicative* error model is to be recommended. Finally, for an analyst who is more concerned with the statistical accuracy of the sample covariance matrices for the volatility parameters used in those statistical models, he/she would rather prefer the surface obtained via *student-t distributed* additive error model.

4.5.2 Monitoring Convergence of the MCMC chains

In this section, we compute the PSRF(V) of the option prices obtained from the σ_{LV} estimates via MCMC. Here, for each error model, we simulate $m = 3$ chains. The m chains are simulated by starting with an over dispersed seed $\sigma_{LV}^{(1)}, \sigma_{LV}^{(2)}, \sigma_{LV}^{(3)}$, where each seed has 89 values for the σ_{LV} parameters. Each chain simulated has a length of 500 after a burn in period of 500 chains. Every kth chain is then saved in a process known as thinning. So, in a chain of 500 estimates, if we choose every 10th estimate, we have a total of 6 estimates in that chain. Thus, for $k = 10$, we have $n = 6$, where n represents the total number of estimates selected from a chain. In our framework, we have tested for $k = 10$ and 100, hence why $n = 6$ and 51 respectively. The estimated parameters through this simulation are then used to construct the local volatility surface via cubic spline interpolation. The surface produced is then used to compute the option prices via finite element method by solving the Dupire PDE. Consequently, the PSRF(V) of the option prices at the selected grid point of the σ_{LV} parameters are computed. Here, we report the PSRF of some of these option prices below. However, similar pattern are observed for the rest of the 84 option prices at those grid points. Also note that, the same analysis can be done for any of the other grid points in which there are no observable prices. Table 4.2 and Table 4.3 show the PSRF for the *additive normally distributed error model* for $n = 6$ and 51 respectively. It can be observed from the two tables that the PSRF gets very close to 1 as n increases from 6 to 51. This is evidence that the

MCMC has converged. Similarly, we compute the PSRF for the option prices for the other error models namely: multiplicative normally distribution error, additive t-distributed error and multiplicative t-distributed error models. These are depicted in Table 4.4 up to Table 4.9.

Table 4.2: PSRF values for the calibrated call price for additive normally distributed error model with $n = 6, m = 3, k = 100$ (using (4.28)).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9234		0.9184	0.9141			
0.25		0.9179		0.9151	0.9145	0.9144		
0.50		0.9155		0.9139	0.9137	0.9140		
0.75		0.9144		0.9140	0.9135	0.9134		
1.0	0.9140		0.9138				0.9131	0.9140
1.5	0.9135		0.9136				0.9139	0.9132
2.0	0.9133		0.9135				0.9132	0.9133

Table 4.3: PSRF values for the calibrated call price for additive normally distributed error model with $n = 51, m = 3, k = 10$ (using (4.28)]).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9983		0.9957	0.9908			
0.25		0.9932		0.9938	0.9913	0.9907		
0.50		0.9904		0.9907	0.9910	0.9902		
0.75		0.9903		0.9902	0.9903	0.9904		
1.0	0.9907		0.9901				0.9906	0.9907
1.5	0.9908		0.9902				0.9903	0.9907
2.0	0.9905		0.9902				0.9903	0.9905

Table 4.4: PSRF values for the calibrated call price for multiplicative normally distributed error model with $n = 6, m = 3, k = 100$ (using (4.28)).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		1.1236		1.1213	0.9489			
0.25		0.9547		0.9593	0.9373	0.9209		
0.50		0.9353		0.9345	0.9243	0.9235		
0.75		0.9261		0.9222	0.9211	0.9216		
1.0	0.9195		0.9216				0.9213	0.9244
1.5	0.9158		0.9183				0.9169	0.9211
2.0	0.9168		0.9162				0.9157	0.9161

Table 4.5: PSRF values for the calibrated call price for multiplicative normally distributed error model with $n = 51, m = 3, k = 10$ (using (4.28)]).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		1.0038		1.0072	1.0364			
0.25		1.0532		1.0588	1.0246	1.0014		
0.50		1.0250		1.0229	1.0077	1.0061		
0.75		1.0107		1.0046	1.0031	1.0043		
1.0	1.0004		1.0037				1.0040	1.0055
1.5	0.9946		0.9989				0.9970	1.0032
2.0	0.9966		0.9958				0.9955	0.9958

Table 4.6: PSRF values for the calibrated call price for additive t distributed error model with $n = 6, m = 3, k = 100$ (using (4.28)).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9170		0.9170	0.9177			
0.25		0.9132		0.9145	0.9152	0.9152		
0.50		0.9133		0.9130	0.9136	0.9140		
0.75		0.9142		0.9131	0.9129	0.9137		
1.0	0.9130		0.9136				0.9151	0.9146
1.5	0.9132		0.9143				0.9129	0.9146
2.0	0.9135		0.9144				0.9129	0.9138

Table 4.7: PSRF values for the calibrated call price for additive t distributed error model with $n = 51, m = 3, k = 10$ (using (4.28)).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9967		0.9964	0.9975			
0.25		0.9907		0.9924	0.9936	0.9936		
0.50		0.9907		0.9903	0.9912	0.9917		
0.75		0.9921		0.9905	0.9902	0.9914		
1.0	0.9903		0.9911				0.9931	0.9924
1.5	0.9906		0.9921				0.9902	0.9925
2.0	0.9912		0.9923				0.9902	0.9915

Table 4.8: PSRF values for the calibrated call price for multiplicative t distributed error model with $n = 6, m = 3, k = 100$ (using (4.28)).

Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9220		0.9177	0.9168			
0.25		0.9172		0.9168	0.9156	0.9149		
0.50		0.9160		0.9162	0.9138	0.9140		
0.75		0.9168		0.9151	0.9141	0.9133		
1.0	0.9139		0.9164				0.9130	0.9134
1.5	0.9132		0.9139				0.9138	0.9129
2.0	0.9131		0.9134				0.9137	0.9130

Table 4.9: PSRF values for the calibrated call price for multiplicative t distributed error model with $n = 51, m = 3, k = 10$ (using (4.28)).

| Maturity	Strike							
	85	90	95	100	110	120	130	150
0.08		0.9922		0.9969	0.9945			
0.25		0.9921		0.9919	0.9924	0.9936		
0.50		0.9924		0.9932	0.9908	0.9909		
0.75		0.9921		0.9929	0.9908	0.9902		
1.0	0.9910		0.9917				0.9909	0.9913
1.5	0.9902		0.9902				0.9913	0.9911
2.0	0.9904		0.9905				0.9917	0.9910

Furthermore, to give additional evidence that the MCMC chains have converged, we compute the $R(k)$ for the four error models studied. We observe that in general, as k increases, precisely $(k = 100, 500, 1000)$, $R(k) \to 0$ for all the models we investigate. This numerically shows the MCMC has converged as well.

4.5.3 Consistency Results for the Local Volatility Surfaces

In this section, we give the numerical results on the consistency of the Bayes' estimators for each of the four statistical model studied namely: normally distributed additive error, student-t distributed additive error, normally distributed multiplicative error, and student-t distributed multiplicative error models. We check for the consistency of the volatility surfaces over four quarterly periods as more and more call option prices are observed in the market.

In general, the surfaces can be seen to be consistent over time as the volatility surfaces remain smooth and stay within a range that prevents arbitrage opportunities in the market. In this context, the volatility values stay within the spread of $[0.1, 0.5]$. This conforms with the volatility values obtained from the market. Figure 4.3, Figure 4.4, Figure 4.5, and Figure 4.6 illustrate the consistency of the surfaces over time.

However, for the first quarter under normally distributed multiplicative error model, we observe high oscillation at lower maturities and strikes. We ascribe this to the fact that we have fewer observations at this quarter due to some options expiring at this period. In addition, this also indicates that the local volatility surface has not yet converged at this period under the normally distributed multiplicative error model while it has shown sign of consistency for the other statistical models.
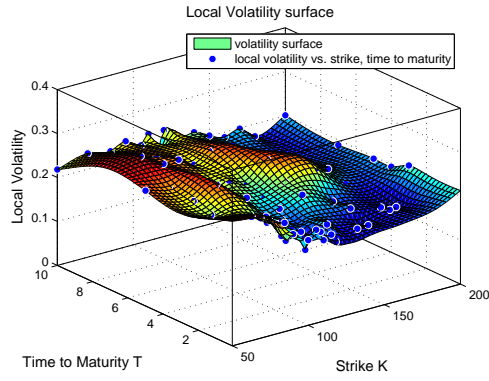
(a) first quarter
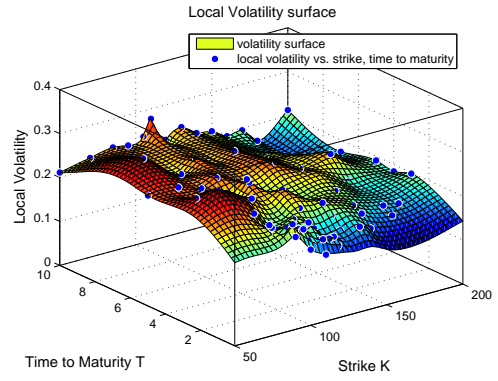
(b) second quarter
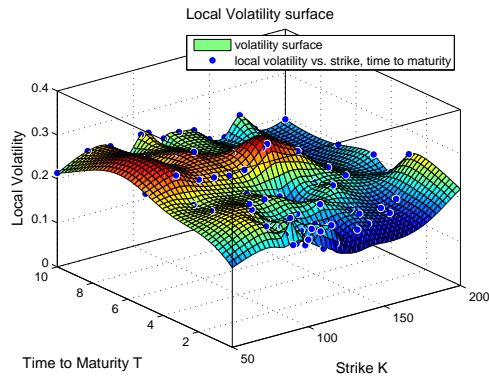
(c) third quarter

(d) fourth quarter

Figure 4.3: Quarterly volatility surfaces for the additive normally distributed error model.

(a) first quarter   (b) second quarter

(c) third quarter   (d) fourth quarter

Figure 4.4: Quarterly volatility surfaces for the multiplicative normally distributed error model.

(a) first quarter

(b) second quarter
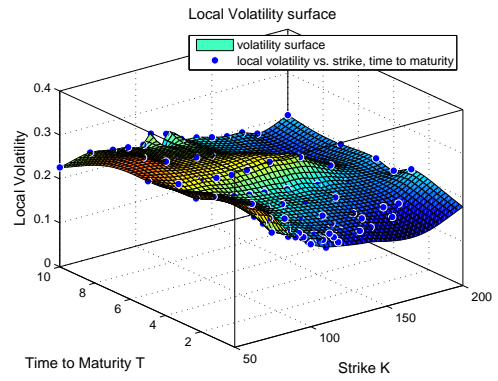
(c) third quarter

(d) fourth quarter

Figure 4.5: Quarterly volatility surfaces for the additive student-t distributed error model.
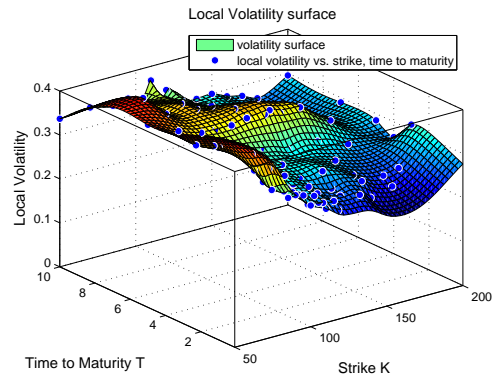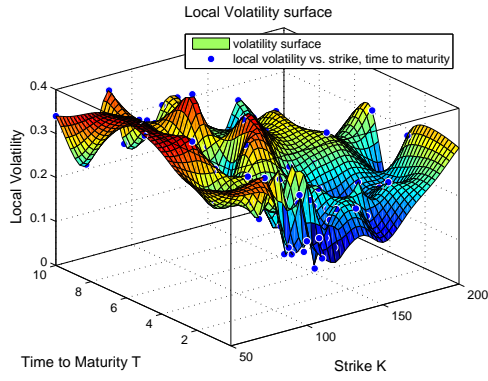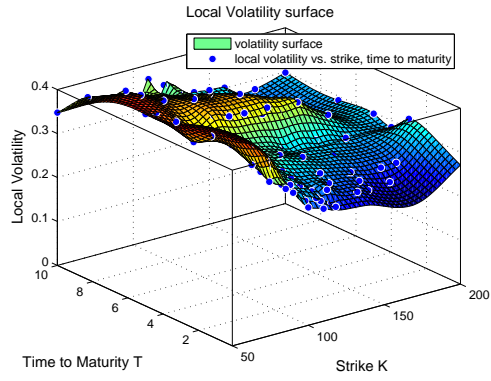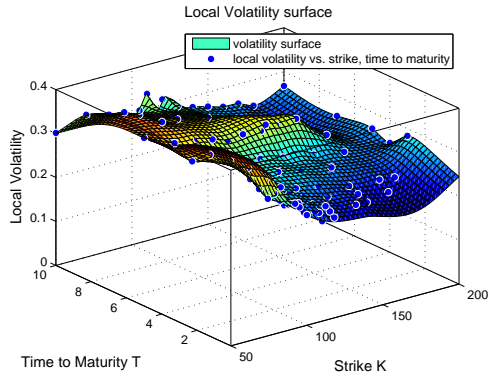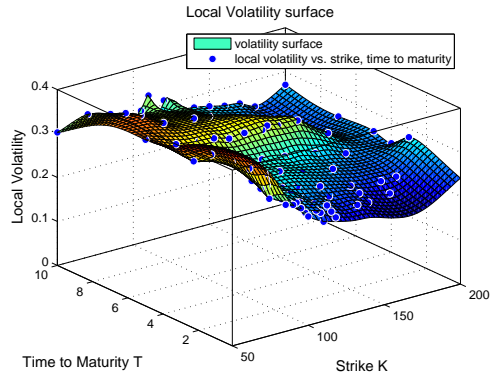
(a) first quarter   (b) second quarter
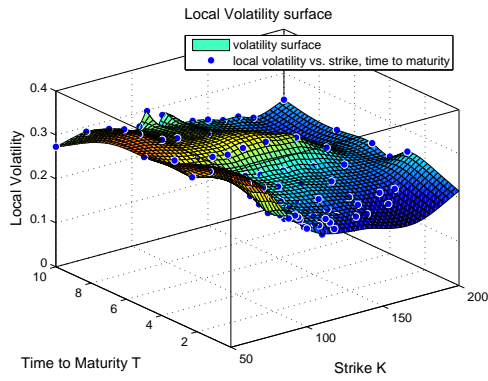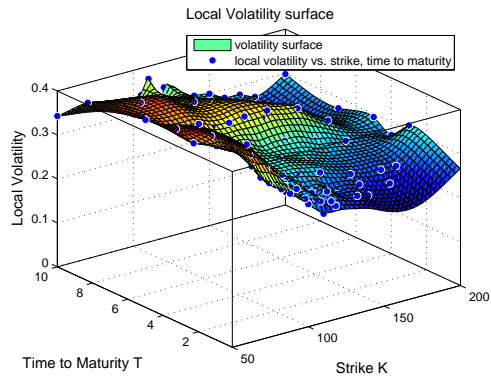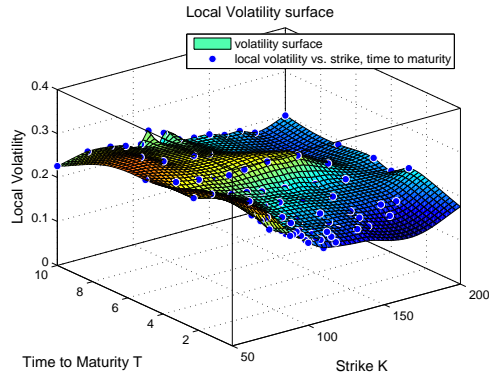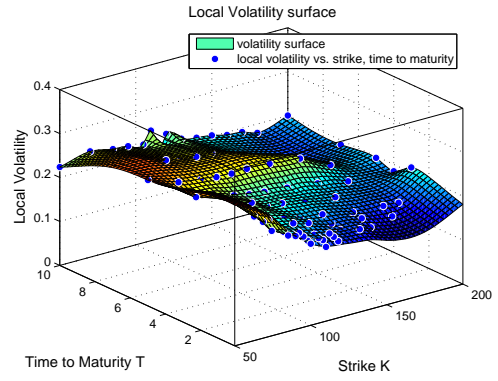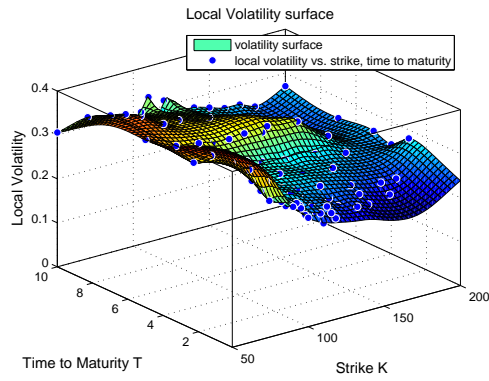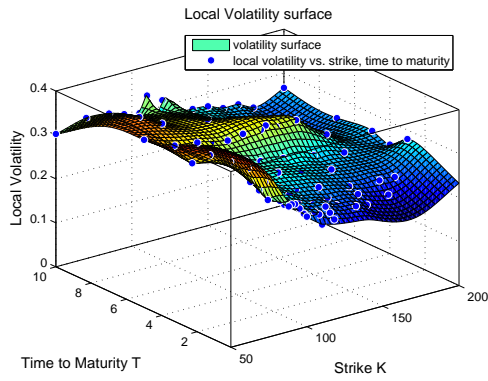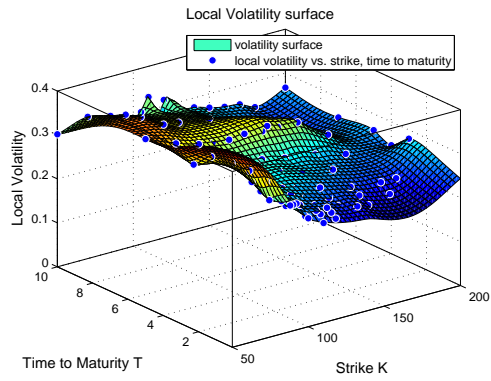
(c) third quarter   (d) fourth quarter

Figure 4.6: Quarterly volatility surfaces for the multiplicative student-t distributed error model.

CHAPTER 5

STOCHASTIC GALERKIN METHOD

Stochastic Galerkin method provides techniques for constructing surrogate models for high-dimensional problems in the parameter space based on low order dimensions. From the perspective of uncertainty quantification, it can be useful to view it as techniques to significantly reduce the *number of deterministic model solutions* required to construct moments associated with a quantity of interest (QoI). To be able to achieve this reduction, one can employ spectral expansions that exploit smoothness or solution techniques with convergence rates faster than the rate, $\frac{1}{\sqrt{M}}$ of Monte Carlo simulation for a moderate parameter expansion. Furthermore, Stochastic Galerkin works well with mutually independent parameters or representation of joint posterior density $\rho_Q(q)$.

The objective of this method is to represent random processes in a manner that exploits the smoothness often exhibited by high-dimensional parameter spaces. This expansion is called *polynomial chaos (PC)* or *spectral expansions*.

In this context, we setup an optimization procedure where we minimize an error functional constrained on solving a parabolic PDE via the Galerkin approach. With this framework, we aim to quantify the option prices as a random output based on the randomness of the volatility parameter.

Let us consider the local volatility PDE again:

$$U_t = \frac{1}{2}\sigma^2(t,x)x^2 U_{xx} - rxU_x, \ \ t > t_0, \ x > 0$$

$$U(t_0, S_0; t_0, x) = \max\{S_0 - x, 0\}, \ x \geq 0,$$

$$U(t_0, S_0; t, 0) = S_0; \lim_{x \to \infty} U(t_0, S_0; t, x) = 0.$$

(5.1)

We then seek to minimize the error functional

$$G(\sigma) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} [U(S_0, 0; T_i, K_{ij}, \sigma) - \tilde{V}_{ij}]^2, \qquad (5.2)$$

where $\tilde{V}_{ij} = \frac{1}{2}(V_{ij}^a + V_{ij}^b)$ is the average of the bid and ask market prices while $U(S_0, 0; T_i, K_{ij}, \sigma)$ is the model price estimated from solving (5.1) under Galerkin approach. The above minimization is done with possibly an addition of a penalty parameter λ to regularize the resultant solution in the parameter space. For this case, we minimize the following functional with respect to σ:

$$G_1(\sigma) = G(\sigma) + \lambda \left\| \left| \nabla \sigma \right| \right\|_2^2.$$

The rest of the chapter is structured as follows: Section 5.1 deals with the representation of a random variable as a linear combination of orthogonal polynomials with random argument. In Section 5.2, we give the weak stochastic formulation for partial differential equations with random inputs. Finally, Section 5.3 uses stochastic Galerkin method to solve the local volatility equation by minimization of the error functional in (5.2). In addition, we use the crude Monte Carlo method to solve the inverse problem and plot the resultant local volatility.

5.1 Polynomial Chaos Expansion

Consider the sequences $\{Q_k(\omega)\}_{k=1}^{\infty}$ of random variables defined on the sample space Ω of the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathbb{P}_k$ be the set of polynomials with arguments Q_i having degrees less than or equal k. Also, let $\hat{\mathbb{P}}_k$ be the set of polynomials in $\mathbb{P}_k$ that are orthogonal to $\mathbb{P}_{k-1}$. $\mathbb{P}_k$ is sometimes termed "Wiener PC of order k when considering Gaussian variable".

Since $Q : \Omega \to \mathbb{R}$, $\mathbb{P}_k$ is a functional; a 2nd order (with finite variance) random variable u can be represented as an infinite expansion

$$
\begin{aligned}
u(\omega) &= u_0 \hat{P}_0 + \sum_{i_1=1}^{\infty} u_{i_1} \hat{P}_1(Q_{i_1}) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} u_{i_1, i_2} \hat{P}_2(Q_{i_1}, Q_{i_2}) \\
&= \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{\infty} \sum_{i_1=1}^{\infty} u_{i_1, i_2, i_3} \hat{P}_3(Q_{i_1}, Q_{i_2}, Q_{i_3}) + \cdots
\end{aligned}
\qquad (5.3)
$$

94

of increasing interaction terms $\hat{P}_i(Q_{i_1}, Q_{i_2}, Q_{i_3}, \ldots, Q_{i_n})$ where $u_{i_1}, u_{i_1,i_2}, u_{i_1,i_2,i_3}, \cdots$ are real coefficients. More compactly, (5.3) can be written as

$$u(Q) = \sum_{k=0}^{\infty} u_k \Psi_k(Q_1, Q_2, \ldots), \tag{5.4}$$

with 1–to–1 correspondence with (5.3).

In practice, we truncate equations (5.3) and (5.4) considering finite set of p random variables with n interaction. For example, truncation of (5.3) at second order interaction gives

$$u(\omega) = u_0 \hat{P}_0 + \sum_{i_1=1}^{p} u_{i_1} \hat{P}_1(Q_{i_1}) + \sum_{i_1=1}^{p} \sum_{i_2=1}^{i_1} u_{i_1,i_2} \hat{P}_2(Q_{i_1}, Q_{i_2}). \tag{5.5}$$

Similarly, truncating (5.4) to k terms yields

$$u(Q) = \sum_{k=0}^{K} u_k \Psi_k(Q_1, Q_2, \ldots, Q_p), \tag{5.6}$$

where $K + 1 = \binom{n+p}{n} = \binom{n+p}{p}$. For clarity, lets consider some cases.

Example 5.1. 1. In essence, for $p = 2, n = 2$ we have

$$\begin{aligned} u(\omega) &= u_0 \hat{P}_0 + \sum_{i_1=1}^{2} u_{i_1} \hat{P}_1(Q_{i_1}) + \sum_{i_1=1}^{2} \sum_{i_2=1}^{i_1} u_{i_1,i_2} \hat{P}_2(Q_{i_1}, Q_{i_2}) = \\ & u_0 \hat{P}_0 + u_1 \hat{P}_1(Q_1) + u_2 \hat{P}_1(Q_2) u_{1,1} \hat{P}_2(Q_1, Q_1) \\ & + u_{2,1} \hat{P}_2(Q_2, Q_1) + u_{2,2} \hat{P}_2(Q_2, Q_2). \end{aligned} \tag{5.7}$$

2. For the case where $n = 2, p = 3$, we get

$$\begin{aligned} u(\omega) &= u_0 \hat{P}_0 + \sum_{i_1=1}^{3} u_{i_1} \hat{P}_1(Q_{i_1}) + \sum_{i_1=1}^{3} \sum_{i_2=1}^{i_1} u_{i_1,i_2} \hat{P}_2(Q_{i_1}, Q_{i_2}) = \\ & u_0 \hat{P}_0 + u_1 \hat{P}_1(Q_1) + u_2 \hat{P}_1(Q_2) u_3 \hat{P}_1(Q_3) \\ & + u_{1,1} \hat{P}_2(Q_1, Q_1) + u_{2,1} \hat{P}_2(Q_2, Q_1) + u_{2,2} \hat{P}_2(Q_2, Q_2) \\ & + u_{3,1} \hat{P}_2(Q_3, Q_1) + u_{3,2} \hat{P}_2(Q_3, Q_2) + u_{3,3} \hat{P}_2(Q_3, Q_3). \end{aligned} \tag{5.8}$$

3. For $n = 3, p = 3$,

$$u(\omega) = u_0 \hat{P}_0 + \sum_{i_1=1}^{3} u_{i_1} \hat{P}_1(Q_{i_1}) + \sum_{i_1=1}^{3} \sum_{i_2=1}^{i_1} u_{i_1,i_2} \hat{P}_2(Q_{i_1}, Q_{i_2})$$

$$+ \sum_{i_1=1}^{3} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} u_{i_1,i_2,i_3} \hat{P}_3(Q_{i_1}, Q_{i_2}, Q_{i_3}) =$$

$$u_0 \hat{P}_0 + u_1 \hat{P}_1(Q_1) + u_2 \hat{P}_1(Q_2) + u_3 \hat{P}_1(Q_3) + u_{1,1} \hat{P}_2(Q_1, Q_1)$$

$$+ u_{2,1} \hat{P}_2(Q_2, Q_1) + u_{2,2} \hat{P}_2(Q_2, Q_2) + u_{3,1} \hat{P}_2(Q_3, Q_1) + u_{3,2} \hat{P}_2(Q_3, Q_2)$$

$$+ u_{3,3} \hat{P}_2(Q_3, Q_3) + u_{1,1,1} \hat{P}_3(Q_1, Q_1, Q_1) + u_{2,1,1} \hat{P}_3(Q_2, Q_1, Q_1)$$

$$+ u_{2,2,1} \hat{P}_3(Q_2, Q_2, Q_1) + u_{2,2,2} \hat{P}_3(Q_2, Q_2, Q_2) + u_{3,1,1} \hat{P}_3(Q_3, Q_1, Q_1)$$

$$+ u_{3,2,1} \hat{P}_3(Q_3, Q_2, Q_1) + u_{3,2,2} \hat{P}_3(Q_3, Q_2, Q_2) + u_{3,3,1} \hat{P}_3(Q_3, Q_3, Q_1)$$

$$+ u_{3,3,2} \hat{P}_3(Q_3, Q_3, Q_2) + u_{3,3,3} \hat{P}_3(Q_3, Q_3, Q_3).$$

$$(5.9)$$

Now, consider a random process $u(t, x, \omega)$ that is a function of random vector $[Q(\omega) = Q_1(\omega), Q_2(\omega), \cdots, Q_n(\omega)] : \Omega \rightarrow \mathbb{R}^n$. By construction, $Q(\omega) \in \Gamma \subset \mathbb{R}^p$ where $\Gamma_i = Q_i(\omega)$ and $\Gamma = \prod_{i=1}^{p} \Gamma_i$ in the image probability space $(\Gamma, \mathcal{B}(\Gamma), \rho_Q(q)dq)$ where $\rho_Q(q)$ is the joint density associated with Q.

For $(t, x) \in [0, T] \times D$, we separate spatial-temporal and random dependencies to obtain

$$u^k(t, x, Q) = \sum_{k=0}^{K} u_k(t, x) \Psi_k(Q), \qquad (5.10)$$

where $u_k(t, x)$ are deterministic coefficients and $\Psi_k(Q)$ are orthogonal polynomials that form the basis for the random component of the solution.

5.1.1   Basis construction for a single random variable

For Q, take $\Psi_k(Q)$ to be 1–D global polynomials that are orthogonal with respect to the density $\rho_Q(q)$ indexed so that $\Psi_0 = 1$. It follows that $\mathbb{E}[\Psi_0(Q)] = 1$ and

$$\mathbb{E}[\Psi_i(Q)\Psi_j(Q)] = \int_\Gamma \Psi_i(q)\Psi_j(q)\rho_Q(q)dq$$
$$= \langle \Psi_i(q), \Psi_j(q) \rangle_\rho = \delta_{ij}\gamma_i, \qquad (5.11)$$

where $\langle \cdot , \cdot \rangle_\rho$ denotes $\mathcal{L}_2$ inner product on the domain Γ with the weight $\rho_Q(q)$. The normalization factor is given by

$$\gamma_i = \mathbb{E}[\Psi_i(Q)\Psi_i(Q)] = \langle \Psi_i(q) , \Psi_j(q) \rangle_\rho .$$

Now, we compute some statistics for the solution of $u^K(t, x, Q)$.

Proposition 5.1.   *1. For fixed t and x, the mean of u^k is given by*

$$
\begin{aligned}
\mathbb{E}[u^K(t, x, Q)] &= \mathbb{E}\left[\sum_{k=0}^{K} u_k(t, x)\Psi_k(Q)\right] \\
&= u_0(t, x)\mathbb{E}\left[\Psi_0(Q)\right] + \sum_{k=1}^{K} u_k(t, x)\mathbb{E}\left[\Psi_k(Q)\right] \quad (5.12) \\
&= u_0(t, x),
\end{aligned}
$$

where $\mathbb{E}[\Psi_0(Q)] = 0$ and $\mathbb{E}[\Psi_k(Q)] = 0$ for $k \geq 1$.

2. Variance of the process can also be calculated as

$$
\begin{aligned}
\mathbb{V}[u^K(t, x, Q)] &= \mathbb{E}\left[\left(u^K(t, x, Q) - \mathbb{E}[u^K(t, x, Q)]\right)^2\right] \\
&= \mathbb{E}\left[\left(u^K(t, x, Q) - u_0(t, x)\right)^2\right] \\
&= \mathbb{E}\left[(u^K(t, x, Q))^2 - 2u_0(t, x)u^K(t, x, Q) + u_0(t, x)^2\right] \\
&= u_0(t, x)^2 - 2u_0(t, x)^2 + \mathbb{E}\left[(u^K(t, x, Q))^2\right] \\
&= \mathbb{E}\left[(u^K(t, x, Q))^2\right] - u_0(t, x)^2 \\
&= \mathbb{E}\left[\left(\sum_{k=0}^{K} u_k(t, x)\Psi_k(Q)\right)^2\right] - u_0(t, x)^2 \\
&= \mathbb{E}\left[2\sum_{k=1}^{K} u_0(t, x)\Psi_0(Q)\Psi_k(Q)\right] + \mathbb{E}\left[\sum_{k=1}^{K} u_k(t, x)\Psi_k(Q)\right]^2 \\
&= \mathbb{E}\left[\left(\sum_{k=1}^{K} u_k(t, x)\Psi_k(Q)\right)^2\right] \\
&= \mathbb{E}\left[\sum_{k=1}^{K} u_k(t, x)^2(\Psi_k(Q))^2\right] + \mathbb{E}\left[2\sum_{1 \leq i < j < K} u_i(t, x)u_j(t, x)\Psi_i(Q)\Psi_j(Q)\right] \\
&= \sum_{k=1}^{K} u_k(t, x)^2\mathbb{E}\left[(\Psi_k(Q))^2\right] + 2\sum_{1 \leq i < j < K} u_i(t, x)u_j(t, x)\mathbb{E}\left[\Psi_i(Q)\Psi_j(Q)\right] \\
&= \sum_{k=1}^{K} u_k(t, x)^2\gamma_k.
\end{aligned}
$$

$$(5.13)$$

Example 5.2 (Hermite polynomials for $Q \sim \mathcal{N}(0,1)$)**.** The pdf for a normally distributed random variable is given by

$$\rho_Q(q) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q^2}{2}\right),$$

which is defined on $\Gamma = \mathbb{R}$. The general form of the Hermite polynomials is given by:

$$H_n(Q) = (-1)^n \exp\left(-\frac{Q^2}{2}\right) \frac{d^n}{dQ^n}\left(\exp\left(-\frac{Q^2}{2}\right)\right). \tag{5.14}$$

The first few Hermite polynomials are: $H_0(Q) = 1$, $H_1(Q) = Q$, $H_2(Q) = Q^2 - 1$, $H_3(Q) = Q^3 - 3Q$, $H_4(Q) = Q^4 - 6Q^2 + 3$.

Hermite polynomials are orthogonal over the real line with respect to Gaussian density. Normalization constant for this set of polynomials is given by

$$\gamma_i = \int_{\mathbb{R}} \Psi^2(q)\rho_Q(q)dq = i\,!. \tag{5.15}$$

We prove the result of (5.15) in the lemma that follows.

Lemma 5.2. *For Hermite polynomials, the normalization constant is given by*

$$\gamma_i = \int_{\mathbb{R}} \Psi^2(q)\rho_Q(q)dq = i\,!. \tag{5.16}$$

Proof. We need to show that

$$\gamma_n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right) H_n^2(q)\rho_Q(q)dq = n\,!. \tag{5.17}$$

As a property of Hermite polynomials [61],

$$H_{n+1}(q) - qH_n(q) + nH_{n-1}(q) = 0; \tag{5.18}$$

indexing with $n - 1$ yields

$$H_n(q) - qH_{n-1}(q) + (n-1)H_{n-2}(q) = 0. \tag{5.19}$$

Multiplying (5.19) by $H_n(q)$ gives

$$H_n^2(q) - qH_{n-1}(q)H_n(q) + (n-1)H_{n-2}(q)H_n(q) = 0. \tag{5.20}$$

Similarly, multiplying (5.18) by $H_{n-1}(q)$ gives

$$H_{n+1}H_{n-1}(q) - qH_n(q)H_{n-1}(q) + nH_{n-1}^2(q) = 0. \tag{5.21}$$

Then, subtracting (5.21) from (5.20) leads to

$$H_n^2(q) - nH_{n-1}^2(q) + (n-1)H_{n-2}(q)H_n(q) - H_{n+1}(q)H_{n-1}(q) = 0. \quad (5.22)$$

(5.22) implies that

$$H_n^2(q) - nH_{n-1}^2(q) = H_{n+1}(q)H_{n-1}(q) - (n-1).H_{n-2}(q)H_n(q). \quad (5.23)$$

Taking the integral over the real line on both sides, we have that

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right) \left(H_n^2(q) - nH_{n-1}^2(q)\right) dq \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right) H_{n+1}(q)H_{n-1}(q)dq \\
&- \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)(n-1)H_{n-2}(q)H_n(q)dq; \\
&\Rightarrow \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)H_n^2(q)dq = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)nH_{n-1}^2(q)dq,
\end{aligned}
\quad (5.24)
$$

where we used the fact that $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)H_m(q)H_n(q) = 0$ for $m \neq n$. Repeating this procedure n-times, we obtain

$$
\begin{aligned}
\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)H_n^2(q)dq &= \frac{n(n-1)(n-2)\cdots 3.2.1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)H_0^2(q)dq \\
&= \frac{n!}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)dq = n!.
\end{aligned}
$$

$$(5.25)$$

Hence, we conclude that

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left(-\frac{q^2}{2}\right)H_n^2(q)dq = n!. \quad (5.26)$$

$\square$

Generally (5.14) is referred to as "probabilistic" Hermite polynomials. The Hermite polynomials have density $\rho_Q(q) = \exp(-q^2)$. With this, we can convert from one Hermite polynomial to another by proper scaling as shown below:

$$
\begin{aligned}
\mathbb{E}[g(q)] &= \int_{\mathbb{R}} g(q)\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q^2}{2}\right)dq \\
&= \int_{\mathbb{R}} g(x\sqrt{2})\frac{1}{\sqrt{2\pi}} \exp\left(-x^2\right)\sqrt{2}dx \\
.&= \int_{\mathbb{R}} g(x\sqrt{2})\frac{1}{\sqrt{\pi}} \exp\left(-x^2\right)dx,
\end{aligned}
\quad (5.27)
$$

given that $x = \frac{q}{\sqrt{2}}$ and $dx = \frac{1}{\sqrt{2}}dq$. For w^r denoting weight of Hermite polynomials and x^r denoting the nodes, we can approximate (5.27) by

$$\mathbb{E}[g(q)] \approx \sum g(x^r \sqrt{2}) \frac{w^r}{\sqrt{\pi}}. \tag{5.28}$$

Hence, the weights and nodes corresponding to (5.14) is $\hat{w}^r = \frac{w^r}{\sqrt{\pi}}$, $q^r = \sqrt{\pi}x^r$.

Example 5.3 (Legendre polynomials for $q \sim \mathcal{U}(-1, 1)$)**.** The first five Legendre polynomials are: $P_0(Q) = 1$, $P_1(Q) = Q$, $P_2(Q) = \frac{3}{2}Q^2 - \frac{1}{2}$, $P_3(Q) = \frac{5}{2}Q^3 - \frac{3}{2}Q$, $P_4(Q) = \frac{35}{8}Q^4 - \frac{15}{4}Q^2 + \frac{3}{8}$. These orthogonal polynomials are defined on the interval $\Gamma = (-1, -1)$ with respect to the density $\rho_Q(q) = \frac{1}{2}$. With this property, they serve as a suitable basis for uniformly distributed random variables on the interval $(-1, -1)$.

Example 5.4. Let $u \sim \mathcal{N}(\mu, \sigma^2)$. Then, the random variable u can be expressed as

$$u = \mu + \sigma Q, \tag{5.29}$$

where $Q \sim \mathcal{N}(0, 1)$. Using Hermite polynomials, u can be written in the form of (5.6) as

$$
\begin{aligned}
u^K &= u_0 H_0(Q) + u_1 H_1(Q) + u_2 H_2(Q) + \cdots \\
&= u_0 + u_1 Q + u_2(Q^2 - 1) + \cdots .
\end{aligned} \tag{5.30}
$$

Comparing (5.29) and (5.30) gives $u_0 = \mu$, $u_1 = \sigma$, $u_k = 0$, $k > 1$.

Example 5.5. Let $u \sim \mathcal{U}(a, b)$. Then, u has mean $\mu = \frac{a+b}{2}$, $\sigma^2 = \frac{(b-a)^2}{12}$. The exact solution is given by $u = \mu + \sqrt{3}\sigma Q$, $Q \sim \mathcal{U}(-1, 1)$. Using Legendre polynomials, we have that $u_0 = \mu$, $u_1 = \sqrt{3}\sigma$, $u_k = 0$, $k > 1$.

Note that we have focused on Hermite polynomials here for the purpose of this thesis. For more information on the classification of hypergeometric orthogonal polynomials (the Askey scheme), we refer interested readers to [89].

5.1.2 Multiple Random Variables

Definition 5.1. A p-tuple $k' = (k_1, k_2, \cdots, k_p) \in \mathbb{N}_0^p$ of non-negative integers is called p-dimensional multi-index with magnitude $|k'| = k_1 + k_2 + \cdots + k_p$ and ordering $j' \leq k' \Leftrightarrow j_i \leq k_i$, for $i = 1, 2, \cdots, p$.

Consider a random vector $Q = [Q_1, Q_2, \cdots, Q_p]$ of mutually independent random variables and let $\{\Psi_k(Q_i)\}_{k=0}^K$ denote univariate basis functions up to degree K in the variables Q_i. Then,

$$\Psi_{i'}(Q) = \Psi_{i_1}(Q_1)\Psi_{i_2}(Q_2)\cdots\Psi_{i_p}(Q_p),$$

for $0 \le |i'| \le K$.

The orthogonality condition is also satisfied by

$$
\begin{aligned}
\mathbb{E}[\Psi_{i'}(Q)\Psi_{j'}(Q)] &= \int_\Gamma \Psi_{i'}(q)\Psi_{j'}(q)\rho_Q(q)dq \\
&= \langle \Psi_{i'}, \Psi_{j'} \rangle_\rho \\
&= \delta_{i',j'}\gamma_i,
\end{aligned}
\tag{5.31}
$$

where $\Gamma = \prod_{i=1}^p \Gamma_k$. Note that $\rho_Q(q)$ is also the product of the densities of each random variable where as $\gamma_i = \mathbb{E}[\Psi_{i'}^2] = \gamma_{i_1}\cdots\gamma_{i_p}$ is the product of univariate normalizing constant. Finally, $\delta_{i',j'} = \delta_{i_1,j_1}\cdots\delta_{i_p,j_p}$ denotes the Kronecker delta up to p variables.

For a random process $u(t, x, Q) : [0, T] \times D \times \Gamma \to \mathbb{R}$ we employ the expansion

$$u^K(t, x, \omega) = \sum_{|k'|=0}^{K} u_{k'}(t, x)\Psi_{k'}(Q), \tag{5.32}$$

with $K + 1 = \binom{n+p}{p}$. Note that $u_{k'}(t, x)$ represents the coefficients of the projection of u^K onto the space of the orthogonal polynomials, $\mathbb{P}_k(Q)$.

Proposition 5.3. *The orthogonality of the basis function can be used to obtain*

$$u_k(t, x) = \frac{1}{\gamma_k}\mathbb{E}\left[u(t, x, Q)\Psi_k\right]. \tag{5.33}$$

Proof. Let the residual be $r = u^K(t, x, Q) - u_k(t, x)\Psi_k(Q)$ when $u^K(t, x, Q)$ is projected onto the space of orthogonal polynomials with basis Ψ_i. Then, $r \perp \Psi_i(Q)$. Thus,

$$
\begin{aligned}
\mathbb{E}\left[r(\Psi_i(Q))\right] = \mathbb{E}\left[\left(u^K(t, x, Q) - u_k(t, x)\Psi_k(Q)\right)\Psi_i(Q)\right] &= 0 \\
\mathbb{E}\left[u^K(t, x, Q)\Psi_k(Q)\right] &= u_k(t, x)\mathbb{E}\left[\Psi_k^2(Q)\right] \\
u_k(t, x) &= \frac{1}{\gamma_k}\mathbb{E}\left[u^K(t, x, Q)\Psi_k(Q)\right].
\end{aligned}
\tag{5.34}
$$

$\square$

As can be seen from (5.33), stochastic Galerkin method deals with the projection of weighted residuals onto a finite dimensional subspace spanned by appropriate basis functions, which provide the constraints required to solve for the deterministic coefficients. Finally, to conclude this section, consider the polynomial chaos expansion in (5.8), we can group the basis functions as shown in Table 5.1.

Table 5.1: Single index, multiple index, and tensored polynomials for $p = 3$

k	$|k'|$	multi-index	Polynomial
0	0	$(0,0,0)$	$\Psi_0(Q_1)\Psi_0(Q_2)\Psi_0(Q_3)$
1	1	$(1,0,0)$	$\Psi_1(Q_1)\Psi_0(Q_2)\Psi_0(Q_3)$
2		$(0,1,0)$	$\Psi_0(Q_1)\Psi_1(Q_2)\Psi_0(Q_3)$
3		$(0,0,1)$	$\Psi_0(Q_1)\Psi_0(Q_2)\Psi_1(Q_3)$
4	2	$(2,0,0)$	$\Psi_2(Q_1)\Psi_0(Q_2)\Psi_0(Q_3)$
5		$(1,1,0)$	$\Psi_1(Q_1)\Psi_1(Q_2)\Psi_0(Q_3)$
6		$(1,0,1)$	$\Psi_1(Q_1)\Psi_0(Q_2)\Psi_1(Q_3)$
7		$(0,2,0)$	$\Psi_0(Q_1)\Psi_2(Q_2)\Psi_0(Q_3)$
8		$(0,1,1)$	$\Psi_0(Q_1)\Psi_1(Q_2)\Psi_1(Q_3)$
9		$(0,0,2)$	$\Psi_0(Q_1)\Psi_0(Q_2)\Psi_2(Q_3)$

5.2 Weak Stochastic Formulation for Partial Differential Equation

In this section, we first set up the *weak stochastic formulation* for a general PDE and subsequently apply the formulation to the local volatility equation.

Definition 5.2 (Weak formulation). Given the PDE

$$
\begin{aligned}
\mathcal{N}(u, Q) &= F(Q), \quad x \in D, \\
B(u, Q) &= G(Q), \quad x \in \delta D,
\end{aligned}
\tag{5.35}
$$

where $\mathcal{N}$ is potentially nonlinear differential operator, $F(Q)$ is a source term, and $B(u, Q)$ and $G(Q)$ are boundary operators. $D \subset \mathbb{R}, \mathbb{R}^2,$ or $\mathbb{R}^3$ and δD denotes the boundary of D. We assume $Q = [Q_1, \ldots, Q_p]$ are mutually independent random variables with range $\Gamma = \mathbb{R}^p$ and joint density $\rho_Q(q)$. The inner product with respect to this density is given by $\langle \cdot, \cdot \rangle_\rho$, while the space of the random variables $L^2_\rho(\Gamma_i)$ has norm $\|g\|_2 = \left(\int_{\Gamma_i} |g(q_i)|^2 \rho_{Q_i}(q_i) dq_i \right)^{\frac{1}{2}}$. We consider solutions

in the space $L^2(0, T; Z)$ where $Z = L^2_\rho(\Gamma) = L^2_{\rho_1}(\Gamma_1) \otimes L^2_{\rho_2}(\Gamma_2) \otimes \cdots \otimes L^2_{\rho_p}(\Gamma_p)$.
Let the quantity of interest be

$$y(x) = \int_\Gamma u(u, q)\rho_Q(q)dq$$

for $x \in D$. The weak formulation of the above PDE involves finding $u \in V$ where
V is a space of test functions that satisfy the boundary conditions of the PDE.
Then, we have that

$$\int_D N(u, Q)S(v)dx = \int_D F(Q)vdx \qquad (5.36)$$

for all $v \in V$, where $S(v)$ is a linear operator and N is a nonlinear operator
constructed using integration by parts.

Example 5.6. Consider the heat equation

$$\alpha\frac{d^2u}{dx^2} = -f(x); \; -1 < x < 1,$$

$$u(-1) = u(1) = 0, \text{ where } Q = \alpha. \qquad (5.37)$$

Here, $\mathcal{N}(u, Q) = \alpha\frac{d^2u}{dx^2}$; $F(Q) = -f(x)$, $B(u, Q) = u$, $G(Q) = 0$, and domain
$D = (-1, 1)$. An appropriate space of test functions is $V = H^1_0(D)$ and the weak
formulation is given by

$$\int_D \alpha\frac{d^2u}{dx^2}v(x)dx = \int_D -f(x)v(x)dx,$$

$$\int_D \alpha\frac{du}{dx}\frac{dv}{dx}dx = \int_D f(x)v(x)dx, \qquad (5.38)$$

which must hold for all $v \in V$. The second line is obtained using integration by
parts where $N(u, Q) = \alpha\frac{du}{dx}$ and $S(v) = \frac{dv}{dx}$.

Definition 5.3. For the differential equations with random inputs, we seek
solutions $u(x, Q) \in V \otimes Z$ which satisfies

$$\int_\Gamma \int_D N(u, q)S(v(x))z(q)\rho_Q(q)dxdq = \int_\Gamma \int_D F(q)v(x)z(q)\rho_Q(q)dxdq \qquad (5.39)$$

for all test functions $v \in V$, and $z \in Z$.

Now, let us revisit the local volatility model in (5.1). We seek solution

$$u(t, x, \sigma) = \sum_{k=0}^{K}\sum_{j=1}^{J} u_{jk}(t)\phi_j(x)\Psi_k(\sigma),$$

where $\{\phi_j(x)\}_{j=1}^J$ and $\{\Psi_k(\sigma)\}_{k=1}^K$ are the basis for the spatial and random spaces such that $V^J = \text{span}\{\phi_j\} \subset V$ and $Z^K = \text{span}\{\Psi_k\} \subset Z$. The choices for ϕ_j are splines, finite elements, or spectral functions. Hence, the domain and space we seek approximate solutions are $D \times \Gamma$ and $V \otimes Z$ respectively.

Following the procedures in (5.39), we have that

$$
\begin{aligned}
& \int_\Gamma \int_D u_t v(x) z(\sigma) \rho_Q(\sigma) dx dq \\
&= \int_\Gamma \int_D \left(\frac{1}{2}\sigma^2(t,x)v(x)x^2 u_{xx} - rv(x)xu_x z(\sigma) \right) \rho_Q(\sigma) dx d\sigma \qquad (5.40) \\
&= \int_\Gamma \int_D \left(\frac{1}{2}\sigma^2(t,x)x^2 u_x v_x - rxu_x v(x) z(\sigma) \right) \rho_Q(\sigma) dx d\sigma.
\end{aligned}
$$

We can approximate the solution further by

$$
\begin{aligned}
& \sum_{m=1}^M \Psi_i(\sigma^m)\rho_Q(\sigma^m)w^m \sum_{k=0}^K \sum_{j=1}^J \frac{du_{jk}}{dt}\Psi_k(\sigma^m) \int_D \phi_j(x)\phi_l(x)dx \\
&= \sum_{m=1}^M \Psi_i(\sigma^m)\rho_Q(\sigma^m)w^m \sum_{k=0}^K \sum_{j=1}^J \frac{1}{2}u_{jk}\Psi_k(\sigma^m) \int_D \phi_j'(x)\phi_l'(x)\sigma^2 x^2 dx \qquad (5.41) \\
&\quad - r\sum_{m=1}^M \Psi_i(\sigma^m)\rho_Q(\sigma^m)w^m \sum_{k=0}^K \sum_{j=1}^J u_{jk}\Psi_k(\sigma^m) \int_D x\phi_j(x)\phi_l(x)dx.
\end{aligned}
$$

5.3 Stochastic Gelarkin for Local Volatility Equation

In this section, we solve the Dupire local volatility equation by writing the option price function of the local volatility equation as a polynomial chaos expansion. Consider the Dupire equation mentioned in (5.1). Using central difference formula for the derivatives, we write the PDE as

$$
U_t - \frac{1}{2}\sigma^2(t,x)x^2 D_2(U) - rxD_1(U) = 0,
$$

where D_1, D_2 are the first and second derivative operators respectively. Furthermore, the θ-method for this problem can be explicitly written as

$$
\frac{U^{n+1} - U^n}{\Delta t} + (1-\theta)\left[-\frac{1}{2}\sigma^2 x^2 D_2(U^n) - rxD_1(U^n) \right] +
$$
$$
\theta\left[-\frac{1}{2}\sigma^2 x^2 D_2(U) - rxD_1(U) \right] = 0. \quad (5.42)
$$

Note that when $\theta = 0$, we have the *explicit method* and when $\theta = 1$, we have the *implicit method.* For $\theta = 1/2$, we have the *trapezoidal rule* in general.

The above formulation can be generalized into the stochastic Galerkin formulation by writing the solution of (5.42) using the generalize polynomial chaos expansion truncated at some integer K. Precisely, we seek solution for the random parametrized PDE

$$U_t(t,x,\omega) - \frac{1}{2}\sigma^2(t,x,\omega)x^2 U_{xx}(t,x,\omega) - rxU_x(t,x,\omega) = 0, \quad t > t_0, \ x > 0, \quad (5.43)$$

such that

$$U(t,x,\omega) = \sum_{i=0}^{K} u_i(t,x)\Psi_i(\xi(\omega)), \quad \sigma(t,x,\omega) = \sum_{j=0}^{K} s_j(t,x)\Psi_j(\xi(\omega)),$$

where $\{\Psi_i(\xi(\omega))\}_{i=1}^{K}$ are the basis for the random space spanned by Ψ_i; that is, $Z^K = \mathrm{span}\{\Psi_i\} \subset Z$. Here, we choose $\xi \equiv [\xi_1, \ldots, \xi_N]^T$ to be independent random variables normally distributed with mean 0 and variance 1. The polynomial chaos expansion of the random variables above has been truncated at K such that for N random variables with P as the highest polynomial interaction we have $K + 1 = \binom{N+P}{P}$. Hence, we can rewrite (5.43) as

$$\sum_{i=0}^{K}(u_i)_t\Psi_i - \frac{1}{2}x^2\left(\sum_{j=0}^{K}s_j(t,x)\Psi_j\right)^2\sum_{i=0}^{K}(u_i)_{xx}\Psi_i - rx\sum_{i=0}^{K}(u_i)_x\Psi_i = 0. \quad (5.44)$$

Furthermore, let $\sigma = (\sigma_1, \ldots, \sigma_L)$ be the parameters of the volatility surface $\sigma(t,x,\omega)$. Using *frequentist statistical approach* as previously studied, let this parameters be normally distribution with implied volatility σ_{imp} as the mean and Σ as the diagonal variance matrix constructed as explained in Section 3.2. Then, we have that

$$s_0(t,x) = \sigma_{imp} \ \text{ and } \ s_1(t,x) = R\xi,$$

where R is obtained from the *Cholesky decomposition* of Σ such that $RR^T = \Sigma$. Therefore, we can write σ as

$$\sigma = \sigma_{imp}(t,x) + R\xi.$$

Moreover, the volatility surface $\sigma(t,x,\omega)$ used in solving the PDE is obtained from the spline interpolation of the $\sigma = (\sigma_1, \ldots, \sigma_L)$. Next, we perform a Galerkin projection on (5.44), which yields

$$(u_k)_t \langle \Psi_k, \Psi_k \rangle - \frac{1}{2}x^2\sigma(t,x)^2\sum_{i=0}^{K}(u_i)_{xx}\langle\Psi_i, \Psi_k\rangle - rx\sum_{i=0}^{K}(u_i)_x\langle\Psi_i, \Psi_k\rangle = 0 \quad (5.45)$$

for $k = 0, 1, \ldots, K$. We then descretize the PCE coefficients using central difference in space and time. Let M be the number of grid points in space. Define $M \times M$ matrices:

$$\mathbf{L}^{(2)} = \frac{1}{2\Delta x^2} \begin{bmatrix} -2 & 1 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & 1 & \\ & & 1 & -2 \end{bmatrix}, \quad \mathbf{L}^{(1)} = \frac{1}{2\Delta x} \begin{bmatrix} 0 & 1 & & & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & 1 & \\ & & -1 & 0 \end{bmatrix},$$

are the second derivative and first derivative operators respectively. Furthermore, define $(K+1) \times (K+1)$ matrix $\mathbf{P}$ with the entries

$$\mathbf{P}_{ik} = \langle \Psi_i, \Psi_k \rangle$$

for $i, k = 0, ..., K$. Also, define the $(K+1) \times (K+1)$ diagonal matrix $\mathbf{E}$ by:

$$\mathbf{E} = \begin{bmatrix} \langle \Psi_0^2 \rangle & & & & \\ & \ddots & & & \\ & & \langle \Psi_j^2 \rangle & & \\ & & & \ddots & \\ & & & & \langle \Psi_K^2 \rangle \end{bmatrix}.$$

Furthermore, define the $M \times M$ diagonal matrices $\mathbf{P}_s^j$ and $\mathbf{P}_x$ by:

$$\mathbf{P}_s^j = \begin{bmatrix} s_j^2(x_1) & & & & \\ & \ddots & & & \\ & & s_j^2(x_m) & & \\ & & & \ddots & \\ & & & & s_j^2(x_M) \end{bmatrix}, \quad \mathbf{P}_x = \begin{bmatrix} x_1 & & & & \\ & \ddots & & & \\ & & x_m & & \\ & & & \ddots & \\ & & & & x_M \end{bmatrix},$$

where $x = (x_1, \ldots, x_M)$ are the points in the spatial grid.

Furthermore, define vectors with lengths $(K+1)M$ such that

$$\mathbf{u}^n = (\mathbf{u}_1^n, \ldots, \mathbf{u}_K^n)^T, \text{ where } \mathbf{u}_i^n = (u_i^n(x_1), \ldots, u_i^n(x_M))^T;$$

$$\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_K)^T, \text{ where } \mathbf{X}_i = (x_1, \ldots, x_M)^T;$$

$$\mathbf{X}^2 = (\mathbf{X}_1^2, \ldots, \mathbf{X}_K^2)^T, \text{ where } \mathbf{X}_i^2 = (x_1^2, \ldots, x_M^2)^T;$$

$$\sigma^n = (\sigma_1^n, \ldots, \sigma_M^n)^T, \text{ where } \sigma_i^n = (\sigma_1^2, \ldots, \sigma_M^2)^T.$$

Next, we define the $M(K+1) \times M(K+1)$ diagonal matrices $\mathbf{P}$ and $\mathbf{P}_x$ by:

$$
\mathbf{P} = \begin{bmatrix} \sigma_1^2 x_1^2 & & & & \\ & \ddots & & & \\ & & \sigma_j^2 x_j^2 & & \\ & & & \ddots & \\ & & & & \sigma_M^2 x_M^2 \end{bmatrix}, \quad \mathbf{P}_x = \begin{bmatrix} x_1 & & & & \\ & \ddots & & & \\ & & x_m & & \\ & & & \ddots & \\ & & & & x_M \end{bmatrix},
$$

where $x = (x_1, \ldots, x_M)$ are the points in the spatial grid. Note that matrices $\mathbf{P}$ and $\mathbf{P}_x$ contains the element wise entries of the vectors $\mathbf{X}, \mathbf{X}^2$, and σ^n. Applying central difference to the partial derivatives of the PCE coefficients in (5.45) and taking the tensor product yields

$$
\left(\mathbf{E} \otimes \mathbf{I}_m\right) \mathbf{u}_t^n - \frac{1}{2}\mathbf{P}\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(2)}\right)\mathbf{u}^n - r\mathbf{P}_x\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(1)}\right)\mathbf{u}^n = 0. \tag{5.46}
$$

Thus, the θ-method can be constructed from (5.46) as

$$
\mathbf{E} \otimes \mathbf{I}_M \left[\frac{\mathbf{u}^{n+1} - \mathbf{u}^n}{\Delta t}\right] + (1-\theta)\left[-\frac{1}{2}\mathbf{P}\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(2)}\right) - r\mathbf{P}_x\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(1)}\right)\right]\mathbf{u}^n
$$

$$
+ \theta\left[-\frac{1}{2}\mathbf{P}\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(2)}\right) - r\mathbf{P}_x\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(1)}\right)\right]\mathbf{u}^{n+1} = 0, \tag{5.47}
$$

where $\mathbf{I}_M$ is $M \times M$ identity matrix. Using *Newton's method* to solve (5.47), we have

$$
F(\mathbf{u}^{n+1}) := \left[\frac{1}{\Delta t}\mathbf{E} \otimes \mathbf{I}_M + \frac{1}{2}\theta\mathbf{P}\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(2)}\right) + \theta r\mathbf{P}_x\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(1)}\right)\right]\mathbf{u}^{n+1}
$$

$$
- \left[\frac{1}{\Delta t}\mathbf{E} \otimes \mathbf{I}_M - \frac{1}{2}(1-\theta)\mathbf{P}\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(2)}\right) - r(1-\theta)\mathbf{P}_x\left(\mathbf{P}_{ik} \otimes \mathbf{L}^{(1)}\right)\right]\mathbf{u}^n
$$

$$
= 0.
$$

$$
\tag{5.48}
$$

5.4   Numerical Results

We solve the system of linear equations in (5.48) of size $(K+1)M \times (K+1)M$ at each time step. Figure 5.1 depicts the local volatility surfaces via the stochastic Galerkin and Monte Carlo methods respectively.

The volatility surface via the stochatic Galerkin approach is obtained by minimizing the error functional in (5.2) while solving (5.48) at each time step. To
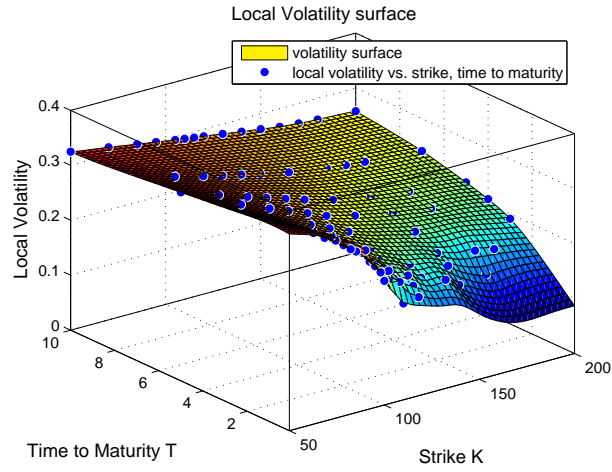
(a) Monte Carlo method with 10000 samples.
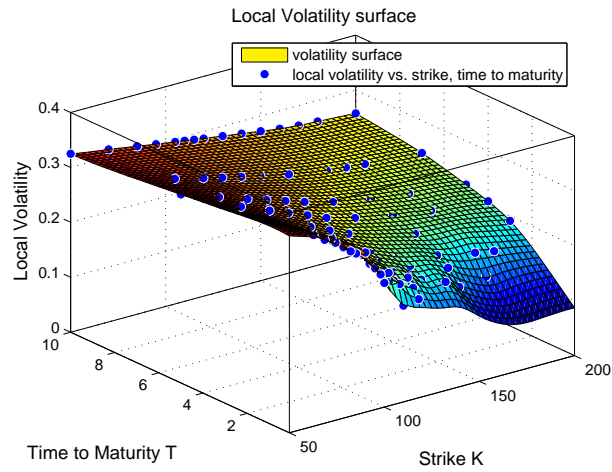
(b) Stochastic Galerkin method by minimization of the
error functional in (5.2).

Figure 5.1: Local volatility surfaces via Monte Carlo method and stochastic
Galerkin methods.

reduce the computational complexity, we have taken the highest order of polynomial interaction to be $P = 3$, and the number of random variables to be $N = 2$, thus making $K = \binom{5}{3} - 1 = 9$. In addition, we have used Hermite polynomials as the orthogonal polynomials since we have worked with normal random variables.

For the Monte Carlo method, we obtain the volatility estimates by solving (5.48) for 2600 and 10000 samples respectively and subsequently taking the expectation of the volatility estimates produced by these samples.

The volatility surfaces from both methods can be used as first hand estimation for the volatility surface. The surfaces also produce volatility values that are within the non-arbitrage volatility spread.

The results of our analysis as depicted in Table 5.2 show that the Monte Carlo method produces higher sums of squared errors (61791) for the 2600 samples compared to the sums of squared errors (61769) computed for the 10000 samples.

Furthermore, the sums of squared errors of the crude Monte Carlo method with 10000 samples produces slightly higher sums of squared errors compared to the stochastic Galerkin method (61764). Thus, the optimization procedure reproduces the market prices better than the crude Monte Carlo method.

In addition, the Monte Carlo procedure with 10000 samples takes significantly more time (6.8534×10^5 seconds) to converge to the sums of squared errors above compared to the stochastic Galerkin method which takes (3.9788×10^3 seconds).

We should also note that the large sums of errors obtained for this procedure is due to the choice of N and P. To obtain more accurate result, a convergence analysis should be performed for higher values of N and P.

Now, comparing the results of all the numerical methods, the normally distributed additive error model has the lowest sums of squared errors as shown in Table 5.2. Furthermore, the results of the Bayesian method demonstrate the lowest sums of squared errors among all the numerical methods with the exception of normally distributed multiplicative error model, where the frequentist

approach performs better.

Table 5.2: Sums of squared errors for all the statistical models and numerical methods used.

Numerical method	Sums of squared errors
Frequentist approach	1701.7
Bayesian (additive normal)	1585.2
Bayesian (additive student-t)	1642.3
Bayesian (multiplicative normal)	2242.9
Bayesian (multiplicative student-t)	1633.7
Stochastic Galerkin	61764
Monte Carlo (2600 samples)	61791
Monte Carlo (10000 samples)	61769

CHAPTER 6

CONCLUSION AND OUTLOOK

In this chapter, we briefly summarize the main results from the previous four chapters. Then, we give some interpretations to these results and finally recommend some areas for further research. In this thesis, we have studied mainly on the calibration of a non-parametric financial model (local volatility) using advanced numerical methods. These methods include frequentist, Bayesian, and stochastic Galerkin. In each of these methods, we have constructed the local volatility surfaces at $t = 0$ that can be used to price options without violating the non-arbitrage principle.

We have made four main contributions to the existing literature on local volatility modeling.

First, we have used the frequentist approach of parameter estimation in constructing a local volatility surface via the sampling distribution of the volatility parameter.

Second, we have shown analytically the consistency of Bayes' estimators for an unknown finite-dimensional (scalar and non-scalar) local volatility parameter under additive and multiplicative normally distributed error models.

Third, we have derived analytically the posterior distributions for the local volatility parameters under the four statistical error models studied. Here, we have also illustrated the convergence of the MCMC estimates using two statistical tests namely, PSRF and autocorrelation.

Finally, we have introduced the use of spectral method to solve the calibration problem. Here, we have estimated the local volatility parameter by minimizing an error functional where the model prices are estimated using stochastic Galerkin approach.

The results from our analysis show that the volatility surfaces via all the three parameter estimation methods we have studied produce surfaces that exhibit the smile effect. Also, the volatility values of all the surfaces are well within the range of values $[0.1, 0.5]$ that would produce non-arbitrage prices in the market.

In addition, our analysis show that the normally distributed additive error model produces volatility surface with the least sums of squared errors (SS) among all the statistical models. However, the normally distributed multiplicative error model produces volatility estimates with the fastest convergence of the MCMC chains among all the statistical error models.

Furthermore, in comparing the sums of squared errors between the frequentist and Bayesian methods, the normally distributed additive error model produces volatility estimates with the lowest SS. However, the frequentist approach produces SS (1701.7) that is lower than the normally distributed multiplicative error model (2242.9) but higher than the other statistical models considered. Thus, the frequentist approach fits the market prices better than the Bayesian method for this case.

The Stochastic Galerkin method has the highest sums of squared errors among all the numerical methods considered. This is expected due to the order of polynomial interaction ($P = 3$) considered. However, in comparison to the Monte Carlo method with 10000 samples used to solve (5.47), we find the stochastic Galerkin to have comparable SS.

An important aspect of this thesis that needs further research is understanding the performance of each of the numerical methods studied. As future work, one can measure the risk of each constructed volatility surface in pricing and hedging path-dependent options. So, instead of using path-independent European calls to calibrate a local volatility model, one would use path-dependent call options.

This can further increase the accuracy of the calibrated parameter due to the improvement in the Bayesian posterior.

Finally, it will also be interesting to analyze the calibration of local volatility function for two or more assets where the assets might be correlated.

REFERENCES

[1] Y. Achdou and O. Pironneau, Volatility smile by multilevel least square, International Journal of Theoretical and Applied Finance, 5(06), pp. 619–643, 2002.

[2] L. Andersen and J. Andreasen, Jump-diffusion processes: Volatility smile fitting and numerical methods for option pricing, Review of Derivatives Research, 4(3), pp. 231–262, 2000.

[3] J. Andreasen and B. N. Huge, Volatility interpolation, Risk, pp. 86–89, 2010.

[4] A. Animoku, Frequentist and Bayesian methods of estimating parameters in a non-performing loan model, WSEAS Transactions on Business and Economics, 15(19), pp. 187–196, 2018.

[5] A. Animoku, Ö. Uğur, and Y. Yolcu-Okur, Modeling and implementation of local volatility surfaces in Bayesian framework, Computational Management Science, 15(2), pp. 239–258, 2018.

[6] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter estimation and inverse problems*, volume 90, Academic Press, 2011.

[7] H. T. Banks, M. Davidian, J. R. Samuels, and K. L. Sutton, An inverse problem statistical methodology summary, in *Mathematical and statistical estimation approaches in epidemiology*, pp. 249–302, Springer, 2009.

[8] H. T. Banks and H. T. Tran, *Mathematical and experimental modeling of physical and biological processes*, CRC Press, 2009.

[9] A. Barron, M. J. Schervish, L. Wasserman, et al., The consistency of posterior distributions in nonparametric problems, The Annals of Statistics, 27(2), pp. 536–561, 1999.

[10] H. Berestycki, J. Busca, and I. Florent, Asymptotics and calibration of local volatility models, Quantitative Finance, 2(1), pp. 61–69, 2002.

[11] J. O. Berger, *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media, 2013.

[12] R. Bhar, C. Chiarella, H. Hung, and W. J. Runggaldier, The volatility of the instantaneous spot interest rate implied by arbitrage pricing–a dynamic Bayesian approach, Automatica, 42(8), pp. 1381–1393, 2006.

[13] P. J. Bickel and J. A. Yahav, Some contributions to the asymptotic theory of Bayes solutions, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 11(4), pp. 257–276, 1969.

[14] F. Black and M. Scholes, The pricing of options and corporate liabilities, Journal of political economy, 81(3), pp. 637–654, 1973.

[15] S. Brooks and G. Roberts, Convergence assessment techniques for Markov chain monte carlo, Statistics and Computing, 8(4), pp. 319–335, 1998.

[16] D. Calvetti and E. Somersalo, *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, Springer, New york, 2007.

[17] M. Cerrato, *The mathematics of derivatives with applications in MATLAB*, John Wiley and Sons, Ltd, 2008.

[18] M. Chance, *An introduction to derivatives and risk management*, Thomson South-Western, 2008.

[19] C. Chiarella, M. Craddock, and N. El-Hassan, The calibration of stock option pricing models using inverse problem methodology, QFRQ Research Papers, UTS Sydney, 2000.

[20] T. F. Coleman, Y. Li, and A. Verma, Reconstructing the unknown local volatility function, Journal of Computational Finance, pp. 77–102, 1999.

[21] P. Constantine, A primer on stochastic Galerkin methods, Lecture Notes, 2007.

[22] R. Cont, Model uncertainty and its impact on the pricing of derivative instruments, Mathematical finance, 16(3), pp. 519–547, 2006.

[23] S. Crépey, Calibration of the local volatility in a generalized Black–Scholes model using Tikhonov regularization, SIAM Journal on Mathematical Analysis, 34(5), pp. 1183–1206, 2003.

[24] S. Crepey, Calibration of the local volatility in a trinomial tree using Tikhonov regularization, SIAM Journal on Mathematical Analysis, 34(5), pp. 1183–1206, 2003.

[25] T. Darsinos and S. Satchell, Bayesian analysis of the Black-Scholes option price, in *Forecasting expected returns in the financial markets*, pp. 117–150, Elsevier, 2007.

[26] T. Darsinos and S. Satchell, Bayesian forecasting of options prices: A natural framework for pooling historical and implied volatility information, in *Forecasting expected returns in the financial markets*, pp. 151–175, Elsevier, 2007.

[27] A. De Cezaro, O. Scherzer, and J. P. Zubelli, Convex regularization of local volatility models from option prices: convergence analysis and rates, Nonlinear Analysis, 75(4), pp. 2398–2415, 2012.

[28] E. Derman and I. Kani, Riding on a smile, RISK magazine, 7(2), pp. 32–39, 1994.

[29] E. Derman and I. Kani, Stochastic implied trees: Arbitrage pricing with stochastic term and strike structure of volatility, International Journal of Theoretical and Applied Finance, 1, pp. 61–110, 1998.

[30] P. Diaconis and D. Freedman, On the consistency of Bayes estimates, The Annals of Statistics, pp. 1–26, 1986.

[31] J. L. Doob, Application of the theory of martingales, The calculation of probabilities and its applications, pp. 23–27, 1949.

[32] B. Dumas, J. Fleming, and R. Whaley, Implied volatility functions: empirical tests, Journal of Finance, 53(6), pp. 2059–2106, 1998.

[33] B. Dupire, Pricing with a smile, RISK magazine, 7(1), pp. 18–20, 1994.

[34] H. Egger and H. W. Engl, Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates, Inverse Problems, 21(3), p. 1027, 2005.

[35] H. Egger, T. Hein, and B. Hofmann, On decoupling of volatility smile and term structure in inverse option pricing, Inverse Problems, 22(4), pp. 1247–1259, 2006.

[36] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, volume 375, Springer Science & Business Media, 1996.

[37] M. R. Fengler, Arbitrage-free smoothing of the implied volatility surface, Quantitative Finance, 9(4), pp. 417–428, 2009.

[38] B. Fitzpatrick, Bayesian analysis in inverse problems, Inverse problems, 7(5), pp. 675–702, 1991.

[39] J. Gatheral, *The volatility surface: A practitional's guide*, John Wiley and Sons, Ltd, 2006.

[40] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*, volume 2, CRC press Boca Raton, FL, 2014.

[41] A. Gelman and D. B. Rubin, Inference from iterative simulation using multiple sequences, Statistical science, pp. 457–472, 1992.

[42] S. Ghosal, A review of consistency and convergence rates of posterior distribution, Proceeding Varanasi Symposium on Bayesian Inference, 1998.

[43] S. Ghosal, A. Van Der Vaart, et al., Convergence rates of posterior distributions for non–iid observations, The Annals of Statistics, 35(1), pp. 192–223, 2007.

[44] M. A. Golberg and H. A. Cho, *Introduction to regression analysis*, Wit Press, 2004.

[45] A. Gupta and C. Reisinger, Robust calibration of financial models using Bayesian estimators, Journal of Computational Finance, 17(4), pp. 3–36, 2014.

[46] A. Gupta, C. Reisinger, and A. Whitley, Model uncertainty and its impact on derivative pricing, Rethinking risk measurement and reporting, 122, 2010.

[47] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward, Managing smile risk, Wilmott Magazine, pp. 84–108, 2002.

[48] M. Hanke and E. Rosler, Computation of local volatilities from regularized Dupire equations, International Journal of Theoretical and Applied Finance, 8(2), pp. 207–222, 2005.

[49] M. Hanke and O. Scherzer, Inverse problem light: Numerical differentiation, American Mathematics Monthly, 108(6), pp. 512–521, 2001.

[50] S. Heston, A closed-form solution for options with stochastic volatility with applications to bond and currency options, Review of Financial Studies, 6(2), pp. 327–343, 1993.

[51] N. Hilber, O. Reichmann, C. Schwab, and C. Winter, *Computational methods for quantitative finance: Finite element methods for derivative pricing*, Springer Science & Business Media, 2013.

[52] B. Hilberink and L. C. Rogers, Optimal capital structure and endogenous default, Finance and Stochastics, 6(2), pp. 237–263, 2002.

[53] J. Hull and A. White, The pricing of options on assets with stochastic volatilities, Journal of Finance, 42(2), pp. 281–300, 1987.

[54] J. C. Hull and S. Basu, *Options, futures, and other derivatives*, Pearson Education India, 2016.

[55] N. Jackson, E. Suli, and S. Howison, Computation of deterministic volatility surfaces, Journal of Computational Finance, 2(2), pp. 5–32, 1999.

[56] E. Jacquier and R. Jarrow, Bayesian analysis of contingent claim model error, Journal of Econometrics, 94(1-2), pp. 145–180, 2000.

[57] E. Jacquier, N. G. Polson, and P. E. Rossi, Bayesian analysis of stochastic volatility models, Journal of Business & Economic Statistics, 20(1), pp. 69–87, 2002.

[58] A. Jobert, A. Platania, and L. Rogers, A Bayesian solution to the equity premium puzzle, Statistical Laboratory, University of Cambridge, 2006.

[59] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*, volume 160, Springer Science & Business Media, 2006.

[60] R. E. Kass, Nonlinear regression analysis and its applications, Journal of the American Statistical Association, 85(410), pp. 594–596, 1990.

[61] R. Koekoek and R. F. Swarttouw, The askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, arXiv preprint math/9602214, 1996.

[62] V. G. Kulkarni, *Modeling and analysis of stochastic systems*, CRC Press, 2016.

[63] R. Lagnado and S. Osher, A technique for calibrating derivative security pricing models: numerical solution for an inverse problem, Journal of Computational Finance, 1(1), pp. 13–25., 1997.

[64] D. Lamberton and B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, Chapman and Hall/CRC: Financial Mathematics Series, 2008.

[65] O. Le Maître and O. M. Knio, *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*, Springer Science & Business Media, 2010.

[66] L. LeCam, On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, University of California Publishing Statistics, 1, pp. 277–330, 1953.

[67] A. A. Markov, Extension of the law of large numbers to dependent quantities (in russian), Izv. Fiz.-Matem. Obsch. Kazan Univ. (2nd Ser), 15(2), pp. 135–156, 1906.

[68] W. Mendenhall, R. J. Beaver, and B. M. Beaver, *Introduction to probability and statistics*, Cengage Learning, 2012.

[69] R. Merton, Option pricing when underlying stock returns are discontinuous, Journal of Financial Economics, 3(1-2), pp. 125–144, 1976.

[70] S. Mikhailov and U. Nögel, *Heston's stochastic volatility model: Implementation, calibration and some extensions*, John Wiley and Sons, 2004.

[71] M. Monoyios, Optimal hedging and parameter uncertainty, IMA Journal of Management Mathematics, 18(4), pp. 331–351, 2007.

[72] N. Moura and N. Silva, *An Introduction to Inverse Problems with Applications*, Berlin, Heidelberg : Springer Berlin Heidelberg : Imprint: Springer, 2013.

[73] M. Musiela and M. Rutkowski, Martingale methods in financial modelling, (2005).

[74] F. D. M. Neto and A. J. da Silva Neto, *An introduction to inverse problems with applications*, Springer Science & Business Media, 2012.

[75] C. S. Ralph, *Uncertainty Quantification: Theory, Implementation and Applications*, Society for Industrial and Applied Mathematics, 2010.

[76] C. S. Ralph, *Uncertainty Quantification: Theory, Implementation and Applications*, SIAM, 2010.

[77] R. Rebonato, *Volatility and correlation: the perfect hedger and the fox*, John Wiley & Sons, 2005.

[78] S. M. Ross, *Introduction to probability models*, Academic press, 2014.

[79] L. Schwartz, On Bayes procedures, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 4(1), pp. 10–26, 1965.

[80] X. Shen and L. Wasserman, Rates of convergence of posterior distributions, Annals of Statistics, pp. 687–714, 2001.

[81] S. E. Shreve, *Stochastic calculus for finance II: Continuous-time models*, volume 11, Springer Science & Business Media, 2004.

[82] D. Sivia and J. Skilling, *Data analysis: a Bayesian tutorial*, OUP Oxford, 2006.

[83] H. Strasser, Consistency of maximum likelihood and Bayes estimates, The Annals of Statistics, pp. 1107–1113, 1981.

[84] A. Tikhonov, Regularization of incorrectly posed problems, Soviet Mathematics, 4, pp. 1624–1627, 1963.

[85] A. Tikhonov and V. Y. Arsenin, *Methods for solving ill-posed problems*, John Wiley and Sons, Inc, 1977.

[86] L. Wasserman, Asymptotic properties of nonparametric Bayesian procedures, Practical nonparametric and semiparametric Bayesian statistics, pp. 293–304, 1998.

[87] D. Xiu, *Numerical methods for stochastic computations: A spectral method approach*, Princeton university press, 2010.

[88] D. Xiu and G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing, 24(2), pp. 619–644, 2002.

[89] D. Xiu and G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing, 24(2), pp. 619–644, 2002.

[90] C. Zhou, A jump-diffusion approach to modeling credit risk and valuing defaultable securities, Working Paper (Federal Reserve:Washington), 1997.

APPENDIX A

PROOFS AND DEFINITIONS

A.1 Proof of Equation (4.23)

For $x, v \in \mathbb{R}$, the following relation holds:

$$v \min(1, x/v) = \min(x, v) = x \min(1, v/x). \tag{A.1}$$

Proof. WLOG let $x \leq v$. This implies that $x/v \leq 1$. and $v/x \geq 1$ Hence,

$$
\begin{aligned}
v \min(1, x/v) &= v(x/v) = x; \\
\min(x, v) &= x; \\
x \min(1, v/x) &= x.1 = x
\end{aligned}
\tag{A.2}
$$

Hence, we conclude that (A.1) is true. □

A.2 Volatility Parameter

First, we restrict σ to a finite dimensional space grid points $K_{\min} = K_1 < K_2 < \cdots < K_m < \ldots < K_M = K_{\max}$ in spatial direction and $T_{\min} = T_1 < T_2 < \ldots < T_n < \cdots < T_N = T_{\max}$ in the temporal direction. We may then represent the volatility parameter by

$$\sigma = (\log \sigma_1, \log \sigma_2, \ldots, \log \sigma_{M+N}),$$

following the ordering convention $\sigma_{m+(n-1)M} = \sigma(k_{m+1}, t_n)$ on the mesh. To determine a complete surface over a specified region $\mathcal{Q}$, we use thin plate cubic

spline to interpolate the discrete volatility values. With this discretization and the prior in (4.36) we have

$$\|\log(\sigma_{LV}) - \log(\sigma_{imp})\|_\kappa^2 = (\log \sigma_{LV} - \log \sigma_{imp})^T C (\log \sigma_{LV} - \log \sigma_{imp}),$$

where C is the inverse covariance matrix induced by the norm.

Second, to calculate the likelihood function in (4.35) for each parameter σ, prices of all the European option values across different strikes and maturities on the given region must be obtained. These prices can be calculated by solving the Dupire's formula (2.13) with appropriate boundary conditions. For the numerical solution of the Dupire's formula we choose the explicit finite difference schemes for its simplicity.

A.3   Discretization of Dupire Local Volatility Equation

In order to solve the Dupire equation on the given parameter space described in Section A.2, we use the Finite Difference Method (FDM). Here, we have used the *forward difference* for the first partial derivatives $\left(\frac{\partial C(T,K)}{\partial T} \text{ and } \frac{\partial C(T,K)}{\partial K}\right)$ and *central difference* for second partial derivative $\left(\frac{1}{2}K^2\frac{\partial^2 C(T,K)}{\partial K^2}\right)$. Applying FDM to the Dupire equation in (2.13) gives

$$
\begin{aligned}
\frac{1}{\Delta T_i}\left(C(T_{i+1}, K_{i,j}) - C(T_i, K_{i,j})\right) = {} & \frac{1}{2}\sigma_{i,j}^2 K_{i,j}^2 \frac{1}{\Delta K^2}\left[C(T_i, K_{i,j-1}) - 2C(T_i, K_{i,j})\right] \\
& + \frac{1}{2}\sigma_{i,j}^2 K_{i,j}^2 \frac{1}{\Delta K^2}C(T_i, K_{i,j+1}) \\
& - rK_{i,j}\frac{1}{\Delta K}\left[C(T_i, K_{i,j+1}) - C(T_i, K_{i,j})\right].
\end{aligned}
$$

$$(A.3)$$

A.4   Induced Inverse Covariance Matrix

Recall that in Chapter 2, Section 4.4 we have defined a functional $\|u\|_\kappa^2$. given by

$$\|u\|_\kappa^2 = (1 - \kappa)\|u\|_2^2 + \kappa \||\nabla u|\|_2^2.$$

Here, $\nabla = \left(\frac{\partial}{\partial T}, \frac{\partial}{\partial K}\right)$ is the gradient operator with $|\nabla u| = \left(\frac{\partial u}{\partial T}\right)^2 + \left(\frac{\partial u}{\partial K}\right)^2$ and $\|\cdot\|_2$ is the standard L_2-norm of the square integrable functions; $\kappa \in (0, 1)$ is a

pre-specified constant.

Let $u : \mathbb{R}^2 \to \mathbb{R}$ be twice differentiable in terms of K and T. Furthermore, let the estimates of u be $\tilde{u}$ at the discretized parameter space corresponding to $L = M \times N$ nodes. Then, we can approximate the first norm $\|u\|_2^2$ by the quadrature rule,

$$\|u\|_{\sim}^2 = \tilde{u}^T \tilde{u} = \tilde{u}^T I \tilde{u},$$

where I is the $L \times L$ identity matrix. For the second norm, consider the integral

$$\||\nabla u|\|_2^2 = \int_T \int_K \left| \frac{\partial u}{\partial K} \right|^2 + \left| \frac{\partial u}{\partial T} \right|^2 dK dT$$

over the rectangle $[K_1, K_2) \times [T_1, T_2)$. Using the notation below

$$u_{j,l} = u(K_m, T_n),$$
$$\triangle_j^K = K_{m+1} - K_m,$$
$$\triangle_l^T = T_{n+1} - T_n,$$

this integral can be approximated by

$$\||\nabla u|\|_{\sim}^2 = \frac{1}{2} \left(\left| \frac{u_{2,1} - u_{1,1}}{\triangle_1^K} \right|^2 + \left| \frac{u_{2,2} - u_{1,2}}{\triangle_1^K} \right|^2 \right) \times \triangle_1^K \triangle_1^T$$
$$+ \frac{1}{2} \left(\left| \frac{u_{1,2} - u_{1,1}}{\triangle_1^T} \right|^2 + \left| \frac{u_{2,2} - u_{2,1}}{\triangle_1^T} \right|^2 \right) \times \triangle_1^K \triangle_1^T.$$

If we represent the parameter space by $[K_{min}, K_{max}] \times [T_{min}, T_{max}]$ with M spatial points $K_{min} = K_1, < \ldots < K_m < \ldots K_M = K_{max}$, and N temporal points $T_{min} = T_1, < \ldots < T_n < \ldots T_N = T_{max}$ chosen in Section A.2, then the approximation to the integral over the whole region is a quadratic function with elements of $\tilde{u}$ given as

$$\||\nabla u|\|_{\sim}^2 = \tilde{u}^T Q \tilde{u},$$

where Q is a positive semi-definite matrix. Thus, the inverse covariance matrix can be represented by

$$C^{-1} = \kappa I + Q.$$

A.5 Expansion of $[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]$.

In Section 4.2, we have proven the consistency of Bayes' estimators for additive and multiplicative error models. As part of the proof, we have bounded the

distribution function of $\sigma_n(V)$. Here, we give more detailed solution on how to bound the quantity $[\phi_n(\alpha_n, V) - \phi_n(\beta_n, V)]$ used as part of that proof for both statistical error models.

In Section 4.2.2, we have derived that

$$
\phi_n(\alpha_n, V) - \phi_n(\beta_n, V) = -2\log(\pi_0(\alpha_n)) + 2\log(\pi_0(\beta_n))
$$
$$
+ \sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]. \quad \text{(A.4)}
$$

Furthermore, we can simplify the expression in the summation term as

$$
\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]
$$
$$
= \sum_{t \in \Upsilon_n} \left[Z_t^2 - 2Z_t \left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right) + \left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]
$$
$$
- \sum_{t \in \Upsilon_n} \left[Z_t^2 - 2Z_t \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right) + \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right]
$$
$$
= 2 \sum_{t \in \Upsilon_n} Z_t \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right) + \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 \right].
$$
$$
\text{(A.5)}
$$

For each $t \in \Upsilon_n$, we can further simply the last term in (A.5) to give

$$
\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t} \right)^2
$$
$$
= \frac{f_t(\alpha_n)^2 + f_t(\sigma^*)^2 - 2f_t(\sigma^*)f_t(\alpha_n)}{\epsilon_t^2} - \frac{f_t(\beta_n)^2 + f_t(\sigma^*)^2 - 2f_t(\sigma^*)f_t(\beta_n)}{\epsilon_t^2}
$$
$$
= \frac{f_t(\alpha_n)^2 - f_t(\beta_n)^2 - 2f_t(\sigma^*)f_t(\alpha_n) + 2f_t(\sigma^*)f_t(\beta_n)}{\epsilon_t^2}
$$
$$
\geq \frac{f_t(\alpha_n)^2 - f_t(\beta_n)^2 - 2f_t(\sigma^*)f_t(\sigma^*) + 2f_t(\sigma^*)f_t(\sigma^*)}{\epsilon_t^2} \quad (\text{since } f_t(\sigma^*) \geq f_t(\alpha_n) \geq f_t(\beta_n))
$$
$$
\geq \frac{f_t(\alpha_n)^2 - f_t(\beta_n)^2}{\epsilon_t^2}
$$
$$
= \frac{(f_t(\alpha_n) - f_t(\beta_n))(f_t(\alpha_n) + f_t(\beta_n))}{\epsilon_t^2}
$$
$$
\geq \frac{(f_t(\alpha_n) - f_t(\beta_n))(f_t(\alpha_n) - f_t(\beta_n))}{\epsilon_t^2}
$$
$$
\geq \left(\frac{f_t(\alpha_n) - f_t(\beta_n)}{\epsilon_t} \right)^2 = \left(\frac{f_t(\beta_n) - f_t(\alpha_n)}{\epsilon_t} \right)^2.
$$
$$
\text{(A.6)}
$$

Substituting the result of (A.6) into (A.5) gives the expression for X_n in (4.11).

In similar fashion, we have also shown in Section 4.2.4 that

$$\phi_n(\alpha_n, V) - \phi_n(\beta_n, V) = -2\log(\pi_0(\alpha_n)) + 2\log(\pi_0(\beta_n)) +$$

$$\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right]. \quad \text{(A.7)}$$

We can further simplify the terms in the summation above as follows:

$$\sum_{t \in \Upsilon_n} \left[\left(Z_t - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(Z_t - \frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right]$$

$$= \sum_{t \in \Upsilon_n} \left[Z_t^2 - 2Z_t \left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right) + \left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 \right]$$

$$- \sum_{t \in \Upsilon_n} \left[Z_t^2 - 2Z_t \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right) + \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right]$$

$$= 2 \sum_{t \in \Upsilon_n} \frac{Z_t}{\epsilon_t} \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{f_t(\beta_n)} - \frac{f_t(\alpha_n) - f_t(\sigma^*)}{f_t(\alpha_n)} \right) \quad \text{(A.8)}$$

$$+ \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right]$$

$$= -2 \sum_{t \in \Upsilon_n} Z_t \left[\frac{f_t(\sigma^*)}{\epsilon_t} \left(\frac{1}{f_t(\beta_n)} - \frac{1}{f_t(\alpha_n)} \right) \right]$$

$$+ \sum_{t \in \Upsilon_n} \left[\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2 \right].$$

For each $t \in \Upsilon_n$, we can further simply the last term in (A.8) to give

$$\left(\frac{f_t(\alpha_n) - f_t(\sigma^*)}{\epsilon_t f_t(\alpha_n)} \right)^2 - \left(\frac{f_t(\beta_n) - f_t(\sigma^*)}{\epsilon_t f_t(\beta_n)} \right)^2$$

$$= \frac{f_t(\alpha_n)^2 + f_t(\sigma^*)^2 - 2f_t(\sigma^*)f_t(\alpha_n)}{\epsilon_t^2 f_t(\alpha_n)^2} - \frac{f_t(\beta_n)^2 + f_t(\sigma^*)^2 - 2f_t(\sigma^*)f_t(\beta_n)}{\epsilon_t^2 f_t(\beta_n)^2}$$

$$= \frac{f_t(\beta_n)^2 f_t(\sigma^*)^2 - f_t(\alpha_n)^2 f_t(\sigma^*)^2 - 2f_t(\sigma^*)f_t(\alpha_n)f_t(\beta_n)^2 + 2f_t(\sigma^*)f_t(\beta_n)f_t(\alpha_n)^2}{\epsilon_t^2 f_t(\alpha_n)^2 f_t(\beta_n)^2}$$

$$\geq \frac{f_t(\beta_n)^2 f_t(\sigma^*)^2 - f_t(\alpha_n)^2 f_t(\sigma^*)^2}{\epsilon_t^2 f_t(\alpha_n)^2 f_t(\beta_n)^2} \quad \text{(since } f_t(\alpha_n) \geq f_t(\beta_n)\text{)}$$

$$= \frac{f_t(\sigma^*)^2}{\epsilon_t^2} \left(\frac{1}{f_t(\beta_n)^2} - \frac{1}{f_t(\alpha_n)^2} \right).$$

$$\text{(A.9)}$$

Substituting the (A.9) into (A.8) gives the desired inequality in (4.18).

A.6 Types of Convergence

When working with stochastic problems, it is often necessary to define sequences of random variables X_n to approximate the exact solution X. When strong convergence can not be established for a random variable, it is often useful to use other forms of convergences like convergence in distribution or probability. Here, we give the definitions of the various form of convergences that exist in the literature.

Definition A.1 (Convergence in distribution)**.** The sequences $\{X_n\}$ converges in distribution to a random variable X, denoted as $X_n \overset{d}{\to} X$ if for all bounded and continuous functions f,

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$$

for $n \to \infty$.

The convergence in distribution holds if and only if there is convergence in the distribution function, F_X, for all continuous points $x \in I_X$, i.e.,

$$X_n \overset{d}{\to} X \longleftrightarrow F_{X_n}(x) \to F_X(x)$$

for $n \to \infty$.

If F_X is continuous, then we have uniform convergence given by

$$\sup_x |F_{X_n}(x) - F_X(x)| \to 0$$

as $n \to \infty$.

Furthermore, let $M_{X_n}(t)$ and $M_X(t)$ be the moment generating functions of the sequence $\{X_n\}$ and random variable X, respectively. Then, the following results hold:

$$\text{If } \lim_{n\to\infty} M_{X_n}(t) = M_X(t) \ \ \forall t \implies X_n \overset{d}{\to} X.$$

Definition A.2 (Almost surely convergence)**.** A sequence $\{X_n\}$ converges a.s to X with probability 1, that is, $X_n \overset{a.s}{\to} X$, if te set ω with $X_n(\omega) \overset{a.s}{\to} X(\omega)$ as $n \to \infty$ with probability 1. In other words, we have almost surely convergence if

$$P(X_n \to X) = P\left(\omega : X_n(\omega) \to X(\omega)\right) = 1.$$

Definition A.3 (Convergence in probability)**.** A sequence $\{X_n\}$ converges in probability to X, $X_n \overset{P}{\to} X$, if $\forall \epsilon > 0$

$$P\left(|X_n - X| > 0\right) \to 0,$$

as $n \to \infty$.

Note that convergence in probability implies convergence in distribution. The converse is true if $X = x$ for constant x. Also, almost surely convergence implies convergence in probability. However, the converse is true for a suitable subsequence $\{X_{n_k}\}$, satisfying $X_n \overset{P}{\to} X \implies X_{n_k} \overset{a.s}{\to} X$.

Definition A.4 (L^p convergence)**.** Let $p > 0$. $\{X_n\}$ converges in L^p or in the pth mean to X, $X_n \overset{L^p}{\to} X$, if $\mathbb{E}\left[|X_n|^p + |X|^p\right] < \infty \ \forall n$, and

$$\mathbb{E}\left[|X_n - X|^p\right] \to 0$$

as $n \to \infty$.

Using the markov inequality, $P\left(|X_n - X| > 0\right) \leq \frac{1}{\epsilon^p}\mathbb{E}\left[|X_n - X|^p\right]$ for $p, \epsilon > 0$, we have that

$$X_n \overset{L^p}{\to} X \implies X_n \overset{P}{\to} X.$$

The converse is not true in general.

APPENDIX B

DATASET

In this appendix, we give the dataset used in constructing the local volatility surfaces in the previous chapters. This data is given in Figure B. The data consists of Bid/Ask implied Black-Scholes volatilities in S&P500, April 1999 taken from [2]. Strikes (K) are in percentage of initial spot and maturities (T) are measured in years. The second column reports approximate option price bid and ask spreads from mid in basis points (1/100 percent) of the spot index value. Volatilities are expressed in percent. Blank cells mean that there are no observations for that particular maturity and strike. The interest rate and dividend yield are 5.59% and 1.14%, respectively.

Table B.1: Numerical examples constant for calibration process in Chapter 4.

Constant	Description	Value
S_0	time 0 asset price	100
r	interest rate	5.59%
d	dividend yield	1.14%
n	# of calibrating options	163
δ	calibration tolerance level (b.p)	3
p	# of calibrated parameters	89
κ	Sobolev norm (LV model)	0.1
m	# of iterations of each MCMC chain	10000
b	length of burn-in	500
dt	time step size	0.01
dK	space step size	5

			Strike (K)																	
T	Bid/Offer		50	70	80	85	90	95	100	105	110	115	120	130	140	150	160	170	180	200
0.08	5	Bid					28.05	25.29	22.17	18.95	15.50									
		Ask					30.13	26.44	23.03	20.10	19.57									
0.25	6	Bid				30.57	28.30	25.95	23.55	21.28	19.23	17.57	15.64							
		Ask				31.75	29.16	26.61	24.13	21.88	23.03	19.03	19.28							
0.50	7	Bid				29.70	27.78	25.96	24.22	22.56	20.98	19.65	18.58	16.22						
		Ask				30.50	28.44	26.52	24.74	23.06	21.52	20.32	19.52	19.04						
0.75	9	Bid			30.96	29.36	27.74	26.15	24.61	23.22	22.01	21.02	20.14	18.50	16.40					
		Ask			31.82	30.08	28.36	26.69	25.11	23.70	22.51	21.56	20.78	19.63	19.22					
1.00	10	Bid			30.60	29.28	27.92	26.53	25.12	23.73	22.43	21.27	20.19	18.27	16.48	12.67				
		Ask			31.40	29.96	28.52	27.07	25.62	24.21	22.91	21.77	20.75	19.12	18.19	18.30				
1.50	11	Bid			30.01	28.93	27.81	26.67	25.53	24.38	23.30	22.37	21.50	19.99	18.68	17.46	16.09	10.74		
		Ask			30.69	29.53	28.35	27.17	25.99	24.82	23.74	22.81	21.96	20.53	19.41	18.64	18.29	18.46		
2.00	12	Bid			29.87	28.90	27.92	26.94	25.97	25.01	24.08	23.24	22.53	21.25	20.11	19.09	18.16	17.25	16.09	
		Ask			30.51	29.46	28.44	27.42	26.41	25.43	24.48	23.64	22.93	21.69	20.63	19.76	19.12	18.72	18.59	
3.00	15	Bid		31.64	30.04	29.24	28.44	27.64	26.84	26.06	25.29	24.56	23.93	22.94	22.04	21.24	20.55	19.97	19.45	18.52
		Ask		32.38	30.64	29.78	28.94	28.10	27.28	26.48	25.69	24.94	24.31	23.32	22.44	21.68	21.05	20.57	20.18	19.73
4.00	17	Bid	34.24	31.69	30.33	29.65	28.98	28.31	27.64	26.99	26.34	25.61	24.97	24.19	23.48	22.85	22.28	21.76	21.31	20.56
		Ask	35.56	32.41	30.91	30.19	29.48	28.77	28.08	27.41	26.74	25.99	25.35	24.55	23.84	23.23	22.68	22.20	21.79	21.18
5.00	20	Bid	33.69	31.60	30.50	29.93	29.38	28.83	28.29	27.75	27.21	26.66	26.14	25.24	24.55	23.98	23.47	22.99	22.55	21.82
		Ask	34.94	32.31	31.08	30.47	29.88	29.31	28.73	28.17	27.61	27.06	26.52	25.60	24.91	24.34	23.83	23.37	22.95	22.28
7.00	42	Bid	32.29	30.97	30.20	29.81	29.42	29.04	28.66	28.29	27.92	27.54	27.17	26.45	25.77	25.31	24.96	24.62	24.32	23.80
		Ask	34.32	32.19	31.22	30.72	30.32	29.88	29.46	29.05	28.64	28.24	27.85	27.09	26.39	25.91	25.54	25.20	24.90	24.40
10.0	75	Bid	31.07	30.50	30.09	29.86	29.63	29.38	29.15	28.91	28.68	28.44	28.20	27.73	27.25	26.78	26.31	25.98	25.77	25.39
		Ask	33.91	32.29	31.61	31.28	30.95	30.64	30.33	30.05	29.76	29.48	29.20	28.67	28.15	27.64	27.13	26.78	26.55	26.15

	Strike (K), % of Spot																	
T	50	70	80	85	90	95	100	105	110	115	120	130	140	150	160	170	180	200
0.08	19.69	19.44	18.29	21.72	21.73	22.11	20.72	19.05	18.20	17.61	17.65	15.75	15.76	14.84	14.31	14.59	15.76	20.27
0.25	19.37	19.04	18.41	20.76	21.50	21.54	20.55	19.18	18.29	17.69	17.48	16.03	15.73	15.04	14.60	14.82	15.83	1966
0.50	18.89	18.44	18.22	19.75	20.83	20.89	20.19	19.18	18.36	17.77	17.38	16.24	15.77	15.26	14.91	15.07	15.87	18.89
0.75	18.41	17.86	17.83	18.99	20.08	20.29	19.79	19.03	18.33	17.79	17.34	16.32	15.82	15.41	15.14	15.24	15.88	18.27
1.00	17.96	17.30	17.39	18.36	19.40	19.73	19.39	18.80	18.22	17.73	17.24	16.33	15.86	15.51	15.28	15.34	15.84	17.78
1.50	17.13	16.21	16.57	17.28	18.18	18.68	18.58	18.22	17.83	17.50	17.11	16.34	15.90	15.65	15.49	15.49	15.77	17.04
2.00	16.50	15.20	15.93	16.52	17.26	17.79	17.90	17.71	17.44	17.20	16.95	16.34	15.94	15.74	15.63	15.60	15.74	16.55
3.00	15.91	13.56	15.26	15.83	16.31	16.67	16.87	16.91	16.82	16.69	16.57	16.31	16.06	15.92	15.85	15.83	15.85	16.07
4.00	15.84	14.41	15.51	15.98	16.26	16.40	16.47	16.48	16.43	16.36	16.29	16.17	16.09	16.10	16.06	16.05	16.06	16.09
5.00	15.98	15.54	16.21	16.37	16.49	16.55	16.54	16.50	16.40	16.28	16.18	16.07	16.04	16.10	16.17	16.21	16.25	16.26
7.00	16.84	16.90	17.02	17.06	17.07	17.05	16.98	16.89	16.78	16.65	16.52	16.28	16.14	16.10	16.15	16.26	16.38	16.58
10.0	18.03	17.99	17.94	17.90	17.85	17.79	17.71	17.63	17.54	17.44	17.33	17.11	16,89	16.71	16.57	16.48	16.45	16.54

Figure B.1: Bid/Ask implied Black-Scholes volatilities in S&P500, April 1999

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Animoku, Abdulwahab

Nationality: Nigeria (NG)

Date and Place of Birth: 07.08.1990, Kogi

Marital Status: Single

Phone: +234 8035 998752

EDUCATION

Degree	Institution	Year of Graduation
PhD	Middle East Technical University	2018
M.S.	Middle East Technical University	2014
B.S.	Middle East Technical University	2012

PROFESSIONAL EXPERIENCE

Year	Place	Enrollment
1 year	Chegg Inc.	Online tutor
2 years	British Culture	English teacher
3 months	American Culture	English teacher

PUBLICATIONS

International Conference Publications

1. A. Animoku, Ö. Ugur, and Y. Yolcu-Okur, Modeling and implementation of local volatility surfaces in Bayesian framework, Computational Management Science, 15(2), pp. 239–258, 2018.

2. A. Animoku, Frequentist and bayesian methods of estimating parameters in a non-performing loan model, WSEAS Transactions on Business and Economics, 15(19), pp. 187–196, 2018.