Statistical Analysis of Volatility Surfaces

Dimitris Lianoudakis

A Thesis in

The Department of Mathematics

and Statistics

Presented in partial fulfillment of the requirements for the degree of Master of

Science (Mathematics) at Concordia University

Montreal, Quebec, Canada

September 2012

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Dimitris Lianoudakis**

Entitled: **Statistical Analysis of Volatility Surfaces**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

<div align="right">Examiner</div>

Dr. A. Stancu

<div align="right">Examiner</div>

Dr. A. Sen

<div align="right">Supervisor</div>

Dr. C. Hyndman

Approved by _____

Chair of Department or Graduate Program Director

_____

Dean of Faculty

Date _____

Abstract

Dimitris Lianoudakis

Option prices can be represented by their corresponding implied volatilities. Implied volatility is dependant on both the strike price and the time to maturity. This dependance creates a mapping known as the implied volatility surface (IVS). The volatility surface is known to practitioners as being synonymous with option prices. These surfaces change dynamically and have distinct features that can be modeled and broken down into a small number of factors. Using time series data of option prices on the S&P500 index, we study the dynamics of the implied volatility surface and deduce a factor model which best represents the surface. We explore the different methods of smoothing the IVS and derive the local volatility function. Using standard dimension reduction techniques and more recent non-linear manifold statistics, we aim to identify and explain these distinct features and show how the surface can be represented by a small number of these prominent factors. A thorough analysis is conducted using principal component analysis (PCA) and common principal component analysis (CPC). We introduce a new form of dimension reduction technique known as principal geodesic analysis (PGA) and give an example. We try to set up a geometric framework for the volatility surface with the aim of applying PGA.

Acknowledgements

I would like to start by expressing my genuine gratitude to my supervisor Dr. Cody Hyndman. His support, guidance and encouragement have greatly facilitated this process. His constant reminder of the light at the end of the tunnel and not to give up was a beacon of hope for me.

I would also like to thank the members of my thesis committee; Dr. Arusharka Sen and Dr. Alina Stancu for their valuable comments. I would like to thank the Institut de finance mathématique de Montréal (IFM2) for funding.

I am thankful for the support I have received from the Department of Mathematics and Statistics over the years, both as an undergrad and as a graduate student. Finally, thank you to my family and friends for having the courage and strength to support me when I had none left.

Contents

List of Figures

List of Tables

Chapter 1

Implied Volatility

1.1 Introduction

The need for volatility modeling has risen dramatically over the past few years and will likely continue to be a high profile activity in the field of quantitative finance. Ever since Black and Scholes [5] and Merton [35] (BSM) introduced their method of fair option pricing in the early 70s, an immense literature in the field of derivative securities has emerged. One such important aspect of this field is volatility modeling. More precisely, there exists a great need to quantify and understand the behavior of volatility in security prices. In a BSM framework the volatility is assumed to be a constant. In other words, there are no underlying assumptions made on any interdependency on this variable, i.e. correlation with the underlying or dependency on other parameters. It is precisely this assumption that make the BSM formula "incorrect". It is obvious that throughout an option's lifetime, its volatility does not remain constant, not even for a day at times. As Rebonato [38] so perfectly put it "A smiley implied volatility is the wrong number to put in the wrong formula to obtain the right price". Because of its ease of calibration, the BSM model has been kept alive by practitioners and many other tractable models have been derived from it.

The main body of this thesis will consist of the following. In the remainder of Chapter 1, we review the theory of option pricing starting with the famed BSM formula. We describe the dynamics and properties of implied volatility (IV) and its importance in understanding risk. We introduce the theory of local volatility and explain its connection with IV and option pricing. Finally, we devote a section of Chapter 1 to our data, explaining the many "cleaning" stages involved before moving on to any analysis. In Chapter 2 we present different techniques used to smooth the implied volatility surface and elaborate on the advantages and disadvantages that each method yields. We also introduce the notion of arbitrage and explore the possibilities of creating an arbitrage-free IVS. The most important part of this thesis is described in Chapters 3 and 4. In Chapter 3 we describe in detail the standard dimension reduction techniques used in the literature, namely principal component analysis, PCA and common principal component analysis, CPC, and compare our results to those obtained by various authors. Our goals are to find and understand the factors that contribute most to the variance of the volatility process, to obtain a lower dimensional parsimonious factor model which can be easily calibrated, and used to price other derivative securities. In Chapter 4 we provide a basic geometric framework in the second part of our dimension reduction analysis. We define the use of a mean and variance on manifold data and contrast the similarities with PCA. We show how the LVS can be represented as a manifold and reason why the use of PCA is not justified anymore. We propose an algorithm for computing PGA on the LVS and suggest future work. In Chapter 5, we end this thesis with our conclusion and closing remarks and discuss the possibility of future research.

1.2 Implied Volatility

What is implied volatility and why is it so important in finance? To even begin answering this question we must first understand where the term "implied" volatility originates from. Volatility, in the general financial context, is used to quantify the risk of a financial instrument over a specific time period. Every type of option has a volatility associated with it. It turns out, that this volatility process plays an important role in the life of the option. We begin by giving a brief overview of the celebrated Black, Scholes and Merton formula used to price these types of equity options.

Consider the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ equipped with a standard Brownian motion process $(B_t)_{0 \leq t \leq T^*}$. Here $\mathcal{P}$ is the real world measure and all information is contained in the filtration $(\mathcal{F})_{0 \leq t \leq T^*}$. The filtration consists of nondecreasing sub-sigma fields $(\mathcal{F}_t)_{t \geq 0}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \ \forall \ 0 \leq s \leq t$. The Black, Scholes and Merton model, is based on a continuous time economy where investors can lend and borrow cash at the risk-free interest rate. No fees are incurred for any transactions and short selling is allowed. We assume a continuously compounded interest rate r, and a stock S_t assumed to be stochastic in nature such that $S = \{S_t : t \geq 0\}$ on $(\Omega, \mathcal{F})$. All trades occur continuously in the interval $[0, T^\star]$, where $T^\star$ is some positive time in the future. Such a process is said to be adapted to a filtration $(\mathcal{F}_t)_{t \geq 0}$ if all S_t are $\mathcal{F}_t$-measurable. S_t is assumed to follow a geometric Brownian motion (GBM) described by the stochastic differential equation:

$$\frac{dS_t}{S_t} = \mu dt + \sigma dB_t \tag{1.1}$$

where $\mu > 0$ is the constant drift process and σ is *assumed* to be the constant volatility process. In a more general context, if both μ and σ depend on the variables (S_t, t), then both processes are said to be adapted to $\mathcal{F}_t$. An important property

3

of equation (1.1Implied Volatilityequation.1.2.1) is that the stock returns are log-normally distributed. That is, the log-returns follow a normal distribution with mean $(\mu - 0.5\sigma^2)(s - t)$ and variance $\sigma^2(s - t)$ for $s \leq t$.

$$\frac{\ln(S_s)}{\ln(S_t)} = \ln(S_s) - \ln(S_t) \tag{1.2}$$

The driving force behind equation (1.1Implied Volatilityequation.1.2.1) is that the BM term is a martingale [24]. In other words, if we assume a constant μ and σ the asset price S_t depends only on the current information and the past has no bearing. Recall the definition of a continuous-time martingale:

Definition 1.1. A stochastic process X_t is called a martingale with respect to the stochastic process Y_t if, for all $t < \infty$:

(a) $\mathbb{E}[|X_t|] < \infty$

(b) $\mathbb{E}[X_t | \{Y_t, t \leq s\}] = X_s, \forall\ s \leq t.$

In particular, for times $s \leq t$

$$\begin{aligned}
\mathbb{E}[B_t | \mathcal{F}_s] &= \mathbb{E}[B_s + (B_t - B_s) | \mathcal{F}_s] \\
&= \mathbb{E}[B_s | \mathcal{F}_s] + \mathbb{E}[(B_t - B_s) | \mathcal{F}_s] \\
&= B_s + \mathbb{E}[(B_t - B_s)] \\
&= B_s,
\end{aligned}$$

where the last expectation is zero since increments of BM are independent and Normally distributed with mean 0 variance $t - s$. This helps us understand the dynamics of S_t in the following way. (b) states that the conditional expectation of an observation at time t+1, given all other observations up to time n, is equal to the observation at time n. In other words, any new information the investor might receive on stock S_t

given that he knows the entire history of the stock up to present time n, is the same as knowing today's current information only. The past has no bearing on knowing the outcome of future events. It is worth noting that the stochastic process S_t is not a martingale, that is $\mathbb{E}[S_t|\mathcal{F}_s] \neq S_s$. Equation (1.1Implied Volatilityequation.1.2.1) is easily solved using Itô's formula. The solution is given by:

$$S_t = S_0 \exp^{(\mu - \frac{1}{2}\sigma^2)t + \sigma B_t} \tag{1.3}$$

where σ^2 measures the risk associated with buying 1 unit of the risky asset S_0.

What exactly causes volatility? It is natural to assume that the volatility of a stock is caused by new information reaching the market. This new information causes investors to react and ultimately changes the value of the stock. This idea however is not supported by research French and Roll [22]. Variance of stock returns between the close of trading on one day and the close of trading on the next day are not proportional. With new information coming in, the change in price does not reflect the time it was received. This leads researchers to suggest that volatility to a certain extent is caused by trading itself.

A call option gives the buyer the right, but not the obligation to purchase an agreed quantity of a particular underlying at a certain time (the expiration date) for a certain price (strike price). A put option gives the seller the right, but not the obligation to sell an agreed quantity of a particular underlying at a certain time (the expiration date) for a certain price (strike price). European style call and put options may only be exercised at maturity, while American style options can be exercised at any time up to the expiration date. For the payoff to be realized, the individual must pay (receive) the option (risk) premium to (from) the counterparty. The buyer

is said to be *long* the option and the counterparty is *short* the option. Whereas the seller is said to be *short* the option and the counterparty is *long* the option. If the option is held to maturity then the counterparty must oblige and deliver the payoff of the agreed conditions stated in the contract entered at the very beginning of the transaction. Very few options; however, are held to maturity. Most investors can close out their positions by entering into the opposing position to sell (buy) the same option. A more mathematical definition of an option is given below.

Definition 1.2. A European contingent claim expiring at time T, is an $\mathcal{F}_T$-measurable positive random variable h.

If $h = (S_T - K)^+$, then h defines a European **call** option with strike price K and exercise date T. The payoff associated with h at time T is:

$$(S_T - K)^+ = \begin{cases} S_T - K, & \text{if } S_T > K \\ 0, & \text{if } S_T \leq K. \end{cases}$$

If $h = (K - S_T)^+$, then h defines a European **put** option with strike price K and exercise date T. The payoff associated with h at time T is:

$$(K - S_T)^+ = \begin{cases} K - S_T, & \text{if } S_T \leq K \\ 0, & \text{if } S_T > K. \end{cases}$$

BSM devised a model for pricing these type of options at $T = 0$. The first step is to set up a portfolio process $H = (H^0, H^1)$ that "replicates" the option payoff. That is we choose the pair (H^0, H^1) such that the value of our portfolio $V_T = S_T^0 H T^0 + S_T^1 H_T^1 = (S_T - K)^+$. If the portfolio is to be self-financing, that is no money is put in or taken out after $t = 0$ and $V_0 = H_0^0 + S_0^0 H_0^1$, then we must have that the price of the option at time zero equal to the value of the portfolio at

6

European Call option with 30 days maturity
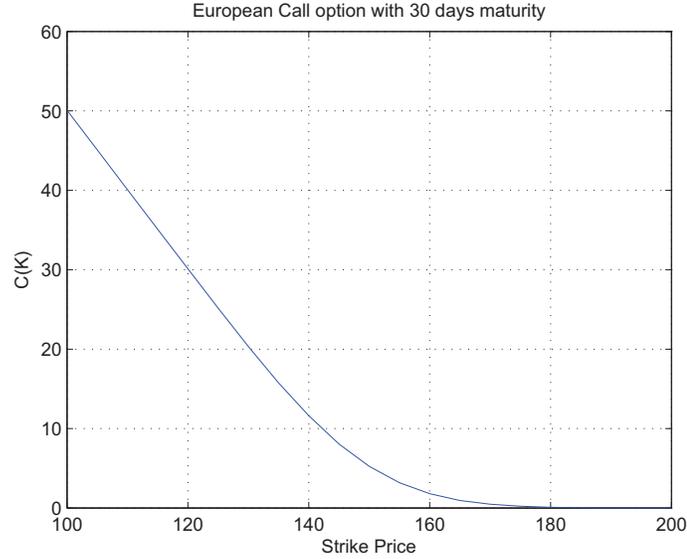
Figure 1.1: Plot of European call option with 30 days to maturity. S_0=\$150 and r=0.01.

time zero. This last statement is derived from the no arbitrage assumption that the investor cannot achieve a risk less profit.

Recall that our modeling framework is $(\Omega, \mathcal{F}_t, P)$ equipped with a BM B_t. We have a bank account S_t^0 and a risky asset S_t^1 with dynamics given by

$$dS_t^0 = rS_t^0 dt, \; S_0^0 = 1$$

$$dS_t^1 = \mu S_t^1 dt + \sigma S_t^1 dB_t, \; S_0 = s_0$$

We want to find a "new" type of measure such that the discounted stock process, $\overline{S}_t^1 = e^{-rt} S_t^1$ is a martingale. This can be achieved by Girsanov's Theorem. Using the same BM and probability space defined earlier let Θ_s be an $\mathcal{F}_t$-adapted square integrable process. Setting $\Theta = \frac{\mu - r}{\sigma}$ denoted as the market price of risk we can define a measure $\mathbb{Q}$ such that the process

$$W_t = B_t + \int_0^t \Theta_u du = B_t + (\frac{\mu - r}{\sigma})t, \tag{1.4}$$

7

is a $(\mathbb{Q}, \mathcal{F}_t)$ BM. Substituting W_t in our asset price equation gives the following dynamics

$$dS_t^1 = rS_t^1 dt + \sigma S_t^1 dW_t \tag{1.5}$$

and it follows that under this new risk-neutral measure, the discounted price

$$d\overline{S}_t^1 = \sigma \overline{S}_t^1 dW_t$$

is a martingale.

The value at time t of a European contingent claim h_T is

$$V_t = \mathbb{E}_Q[e^{-r(T-t)} h_T | \mathcal{F}_t]. \tag{1.6}$$

The solution to equation (1.6Implied Volatilityequation.1.2.6) when $h_T = (S_T - K)^+$ can be found in [43], and is known as the celebrated BSM Call function

$$C_{BS}(K, S(t), r, q, \tau, \sigma) = S_t N(d_1) - N(d_2) K e^{-(r-q)\tau}, \tag{1.7}$$

where

$$d_1 = \frac{\ln(S/K) + ((r-q) + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$
$$d_2 = d_1 - \sigma\sqrt{T-t}.$$

$N(\cdot)$ is the cumulative distribution function of the standard Normal distribution, r is the risk free interest rate compounded continuously, q the continuous dividend yield, $\tau = T - t$ is the time to maturity, and σ, the stock option's volatility (assumed to be constant throughout its lifetime). We can reduce the dimensions of the BSM equation by inducing a transformation of variables. The BSM implied total variance

w is defined as:

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

and the log-moneyness y defined as

$$y = \log\left(\frac{K}{F_t}\right)$$

t where $F_t = S_0 e^{(r-q)(T-t)}$ is the forward or futures price of the stock at time t. In terms of these two dimensionless variables and the futures price F_t, the reduced BSM equation can be written as:

$$
\begin{aligned}
C_{BSM}(F_T, y, w) &= F_T\{N(d_1) - e^y N(d_2)\} \\
&= F_T\left\{N\left(-\frac{y}{\sqrt{w}} + \frac{\sqrt{w}}{2}\right) - e^y N\left(-\frac{y}{\sqrt{w}} - \frac{\sqrt{w}}{2}\right)\right\}. \quad (1.8)
\end{aligned}
$$

The futures moneyness metric is defined as:

$$\kappa = \frac{K}{F_t}$$

where F_t is defined as above. A call option is in-the-money (ITM) if $\kappa > 1$, out-of-the-money (OTM) if $\kappa < 1$ and at-the-money (ATM) if $\kappa \approx 1$. For puts we just need to reverse the equalities.

The only unknown parameter in equation (1.7Implied Volatilityequation.1.2.7) is the volatility parameter, σ. BSM postulated that the price of a call option depends only on the variables t and S and assume that all other variables are known. In practice this is not the case as observed by ever changing market conditions, i.e. for fixed strike price K and time to maturity τ the price of a call option changes from day to day. Thus σ is no longer a fixed value but a varying parameter. In a perfect world we would have that market call prices agree with call prices derived from the

BSM equation.

For a fixed T, call prices are monotone and convex in the K direction. This is the hockey stick shape we see in Figure 1.1. The monotonicity and convexity property of these plain vanilla options enables us to circumvent the nonlinearity of their nature and numerically solve for the IV. Denote observed market call options by $C_t^{MKT}(K,T)$, then the BSM implied volatility $\sigma(K,T)$ is the parameter σ which equates observed prices to those obtained from the BSM equation, or:

$$C_{BS}(K, S(t), r, q, \tau, \sigma_t^{BS}(K,T)) = C_t^{MKT}(K,T) \qquad (1.9)$$

for all $\sigma_t^{BS}(K,T) \geq 0$.

The implied volatility $\sigma(K,T)$, is a convex function of the strike price. It displays a pronounced curvature referred to by many as the volatility smile as seen in Figure 1.2. This contradicts the BSM model assumption that $\sigma(:,T)$ should be a constant $\forall$ $t \in [0,T]$. Taking note of the above, IV can then be viewed as a mapping from time, strike, and time to maturity to the positive reals numbers. This map is known as the *Implied Volatility Surface*, denoted as IVS form here on, such that:

$$\widehat{\sigma}_t : (t, K, T) \mapsto \widehat{\sigma}_t(K,T) \,\forall\, t \geq 0.$$

It can be noted that if σ is no longer a deterministic parameter, but rather a function of time and the stock price, then S_t no longer follows a GBM, and the BSM equation is no longer valid. The value of σ which equates the market prices to those found using BSM is found by reverse engineering equation (1.9Implied Volatilityequation.1.2.9) for the diffusion parameter $\sigma_t^{BS}(K,T)$. Exploiting the fact that the BSM price is monotone in σ enables us to find a unique solution $\hat{\sigma} > 0$ such that equation (1.9Implied Volatilityequation.1.2.9) is satisfied and uniqueness is
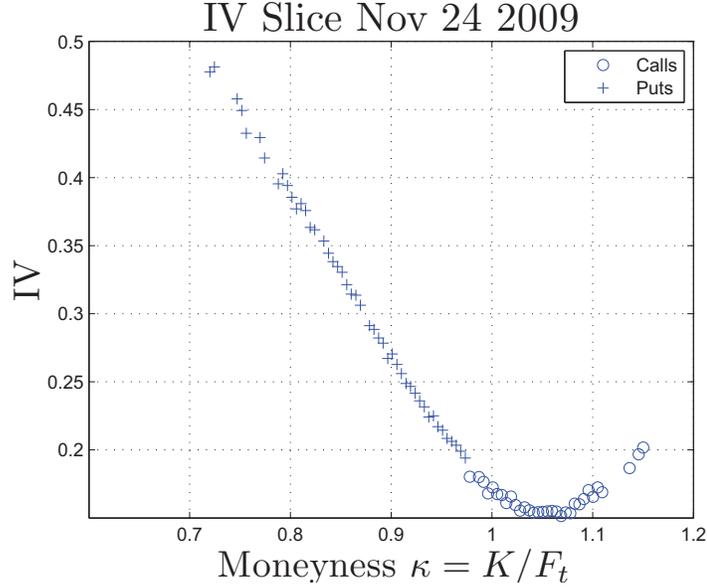
Figure 1.2: Implied Volatility from Call and Put Option prices. Lower axis has been scaled by the moneyness metric $\kappa = \frac{K}{S_t}$. Only OTM call and put prices have been used. Calls are 'o'-shaped and puts are '+'-shaped.

guaranteed by the monotonicity of the call price function. A simple Newton-Raphson algorithm can be used to find the value of $\hat{\sigma}$.

The IVS is scaled across both the strike and time axes. A moneyness (κ) and time to maturity (τ) metric are used to reduce the overall dimension of the BSM equation. Since these are European style options, they can only be exercised at expiry, and so the choice of the the futures moneyness metric is convenient as it also incorporates the risk neutral drift μ in its measure. Stock price moneyness can also be used and is defined as:

$$\kappa = \frac{K}{S_t}.$$

Put-call parity states that under the BSM assumptions, IVs for both put and call options must be equal. If this assumption is violated, then simple arbitrage exists. Unfortunately this is often the case, and it is observed that IV for puts are higher than IV for calls. Figure 1.2 will be useful in the following explanations. As markets

rise, there are always investors who are willing to sell. Consequently, investors are willing to pay a lower premium (lower IV) for OTM calls since they believe that there is a small chance the stock will fluctuate in price. As markets fall; however, and potential losses can occur, people are willing to pay a higher premium to own OTM puts. These strategies and others are used by hedge funds as a sort of insurance against such a downward movement. We also observe form Figure 1.2 that there are more OTM put IVs than there are OTM call IVs, which translates to more OTM put options than OTM call options. Hence, investors awareness of potential losses as reflected by the smirk structure of the IV curve have increased and thus, are willing to pay more to "insure" against such events.

Implied volatility data is often "contaminated" in such a way that one can find arbitrage opportunities across strikes and maturities. Even when the initial data is arbitrage-free, smoothing the IVS via parametric or nonparametric techniques can generate surfaces that are not always arbitrage-free. Estimating the IVS under arbitrage-free conditions is a tricky task. The form of the IVS is specific at each point in time, it requires estimating in a high-dimensional space, while taking into account the effect of the variables, such as the spot price, interest rate, strike price, etc. Methods for computing an arbitrage-free IV surface will be discussed in more detail in Chapter 2.

Most trading is done around the ATM region as can be observed by the clustering of points in Figures 1.2 and 1.3. Our option data consists of end-of-day prices where noise and potential outliers can be found. By using standard techniques in data analysis, we have "cleaned" the data of any potential arbitrage violations and outliers so as to not comprise the integrity of the surface. The topic of data management will be further examined in detail in Chapter 2.
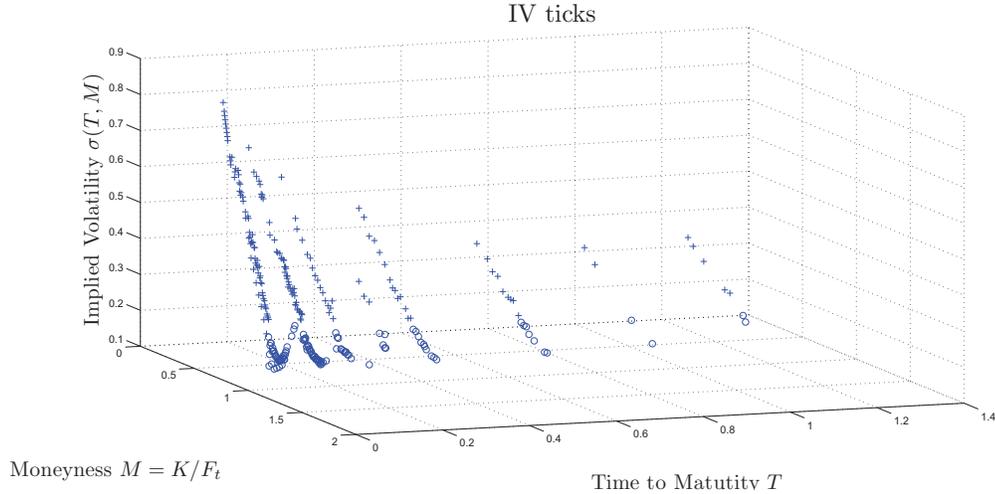
Figure 1.3: November 11, 2009 implied volatility ticks for call and put options. Left axis denotes moneyness define as, $\kappa = K/S_t e^{(r-q)\tau}$ on the interval [0.5,2] and right axis denotes time to maturity defined as , $\tau = T - t$ on the interval [0,1] measured in years.

Below, we state some useful definitions of the standard terminology used in the IVS literature.

- **Volatility smile/skew** - The dependence of IV on the strike price for a fixed maturity

- **Volatility term structure** - The dependence of IV on the time to maturity for a fixed level of moneyness, usually for ATM options.

- **Implied Volatility Surface** - The dependence of IV on both the strike price and time to maturity

The IVS fluctuates between the bid-ask spread. The wider the spread, the more pronounced the smile may be. Another important property is its degenerative behavior across time to maturity. The further along one goes across the time to maturity axis, the lower the IV at that value, and the shallower the surface. This can be viewed as investors having a low risk appetite for options maturing far into the future than

13

those closer to the present. The "pearl" like strings of the IVS seem to grow thinner in number the further away in time you go. This characteristic is well displayed in Figure 1.3. One of the reasons is that some contracts are not traded as much as others, and so the longer the maturity the lower the volatility. The smile achieves its minimum at the ATM to near ITM region; see Figures 1.2 and 1.3. The upside of the BSM model is that IV is taken at face value. More often than not, IV is seen as the market's expectation of average volatility throughout the life cycle of the option. This is to our favor because if the market believes it is in a high or low volatile state, it will be reflected in the IVS. Traders can quote option prices in terms of IV rather than the underlying and make a market for IV (like the VIX[1] index).

The underlying S_t assumed to follow a log-normal distribution cannot account for the probability of large downward movements, therefore falsifying the BSM formula. The BSM formula is then only employed as a means of computational check amongst traders. The IVS is very strike dependent and time to maturity dependent, thus for the same set of options, a different IVS will be created everyday, making it difficult for static hedging portfolios. A typical picture of the IVS is displayed in Figure 1.4 for November 11, 2009.

There is strong evidence to suggest that IVs appear to be mean-reverting in nature [9], "Shocks" along the IVS are indeed highly correlated, and there is evidence to suggest that the surface can be reduced into a small number of principal factors (components). More of this will be discussed in the dimension reduction part of this thesis. It is precisely these fluctuations which we focus on and try to determine how they can be quantified in a coherent model.

---

[1]The CBOE Volatility Index (VIX) is a key measure of market expectations of near-term volatility conveyed by S&P500 stock index option prices.
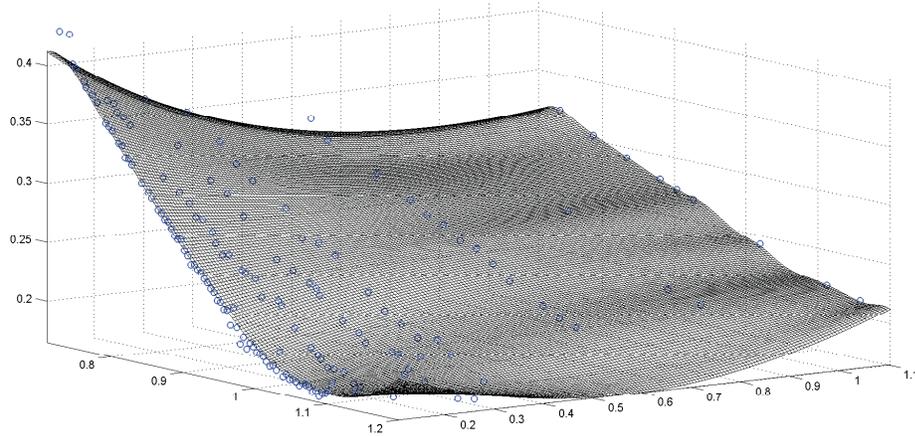
Figure 1.4: Implied Volatility Surface of November 11, 2009. Left axis denotes moneyness define as, κ and right axis denotes time to maturity defined as , $\tau = T - t$.

1.2.1 Interpreting the Smile

For a fixed time to maturity, we get a one-dimensional representation of the IV smile or "slice" in the τ direction. This slice often called the smile or smirk and is evident of post 1987 crash where a typical IV slice exhibited a U-shape with its minimum at or near the money ($\kappa \approx 1$). Since then the risk awareness of investors has changed dramatically and the slice is more downward sloping at and near the money (.95 $\leq \kappa \leq 1.05$) and tends to curve upwards for OTM strikes ($\kappa \leq 1$).

The smile tells us that there is a premium charged for OTM put options and ITM call options above their computed BSM price with the ATM IV. In other words, the market is "high pricing" these options as if the log-normal model of the underlying fails to capture probabilities of large downward movements. This is what we see in Figure 1.5. Slope bounds can be obtained for a volatility curve $\sigma(K)$, by noting that calls must be decreasing in the K direction, or else arbitrage opportunities exist. More of this will be tackled in the arbitrage section of this thesis.

	(a) IV Smile	(b) Smoothed IV Smile
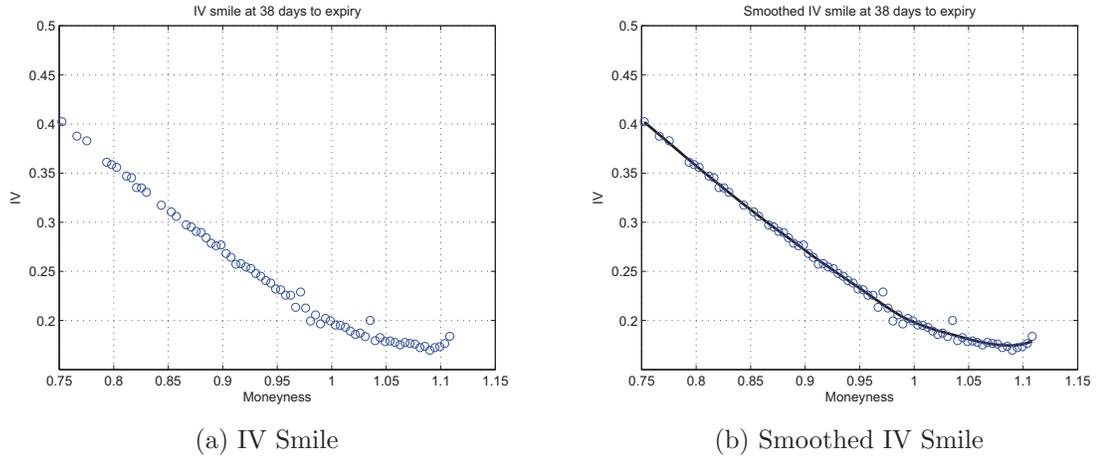
Figure 1.5: Left Panel: Implied volatility smile for November 11, 2009. Left axis denotes IV ticks and right axis denotes moneyness. Right Panel: Implied volatility smile for November 11, 2009. Smile has been smoothed using a local linear polynomial with localized bandwidth h_κ. Left axis denotes IV ticks and right axis denotes moneyness define as, $\kappa = K/S_t e^{(r-q)\tau}$

One can also observe the relative term structure of IVs and their associated degenerate nature. Figure 1.6 displays the term structure for various slices of the IVS for a ranging level of moneyness; $\kappa = 0.75$ top line, $\kappa = 1$, middle line, and $\kappa = 1.10$, bottom line. We notice that the ATM IV and OTM calls (ITM puts) exhibits a slightly increasing slope, while OTM puts (ITM calls) display a decreasing term structure. This indeed corresponds to the more shallow smile of the IVS for longer term maturities.

1.3 Local Volatility

The following section will be devoted to the Dupire [14] equation and the theory of *Local Volatility* (LV). Implied volatility predictions or forecasts do not depend on historical prices or on historical volatilities. There are many reasons why a "smile" exists; Variations in the bid-ask spread, market liquidity, stochastic volatility, just to name a few. The IVS can then be seen as a global measure of volatility. It was
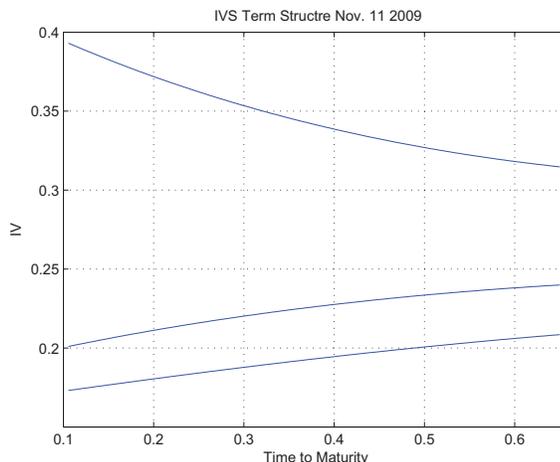
Figure 1.6: Term Structure of the IVS for November 11, 2009. With relative levels of moneyness: $\kappa = 0.75$, top line, $\kappa = 1.00$, middle line, and $\kappa = 1.10$, bottom line. Time to maturity is measured in years.

initially understood that the risk-neutral density (RND) could only be derived from the market prices of European options. That notion changed when Dupire, and Derman and Kani [12] independently showed that under the risk neutral measure, there exists a unique diffusion process that is consistent with the distribution of option prices. This unique *state-dependent* diffusion process is known as the *Local Volatility function* and is denoted by $\sigma_{LV}(S,t)$.

Local volatility is used by practitioners as an efficient way to price exotic options consistently with given prices of vanilla options. The diffusion coefficient $\sigma_{LV}(S,t)$ can be derived from the classical *Forward Kolmogorov* or *Fokker-Planck* PDE. The derivation will require us to work backwards. We are going from a time, strike and time to maturity dependant implied volatility, $\sigma_{BSM}(t,K,T)$ to a time and spot dependant volatility known as the *Local Volatility Surface* (LVS), $\sigma_{LV}(S_t,t)$. The local volatility function can be seen as the instantaneous volatility for a market level K at some future date T^*. Let us begin with a description of the model.

We consider the usual space $(\Omega, \mathcal{F}, \mathcal{P})$, a filtration $\mathcal{F}_t$, and a Brownian motion process W_t for $0 \leq t \leq T^*$ such that the dynamics of S_t can de described by the SDE

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma(S_t, t, \cdot)dW_t.$$

where $\mu(S_t, t)$ is the instantaneous drift and $\sigma(S_t, t, \cdot)$ follows some $\mathcal{F}_t$-measurable process depending on S_t, t and possibly some other variables.

For the risk neutral measure $\mathbb{Q}$ to exist we assume the absence of arbitrage and so S_t is a martingale. Our state space where the set of all European plain vanilla call options exist can be defined as

$$\mathfrak{G} = \{C_t(K, T) : K \geq 0, 0 \leq T \leq T^*\}.$$

Fengler [17] shows that the local variance function $\sigma^2_{K,T}(S_t, t)$ is defined as the expectation of the squared instantaneous volatility under the risk-neutral measure conditional on $S_t = K$ and time t on $\mathcal{F}_t$, that is

$$\sigma^2_{K,T}(S_t, t) = \mathbb{E}_{\mathbb{Q}}[\sigma^2(S_T, T, \cdot)|S_T = K, \mathcal{F}_t]. \tag{1.10}$$

The local volatility is then

$$\sigma_{K,T}(S_t, t) = \sqrt{\sigma^2_{K,T}(S_t, t)}. \tag{1.11}$$

Indeed if the instantaneous volatility is deterministic (non-stochastic) in spot and time only, namely, $\sigma(S_t, t, \cdot) = \sigma(S_t, t)$, then the notion of instantaneous volatility

and local volatility is one and the same, i.e

$$\sigma^2_{K,T}(S_t, t) = \mathbb{E}_{\mathbb{Q}}[\sigma^2(S_T, T, \cdot)|S_T = K, \mathcal{F}_t]$$

$$= \mathbb{E}_{\mathbb{Q}}[\sigma^2(S_T, T)|S_T = K, \mathcal{F}_t]$$

$$= \sigma^2(K, T).$$

For a detailed summary of the above equality see Gatheral's [23] chapter on local volatility. The value of the local volatility function can be observed directly from the value of S, where as implied volatility is an unobservable variable. Equations (1.10Local Volatilityequation.1.3.10) and (1.11Local Volatilityequation.1.3.11) have an intuitive meaning. As Gatheral states, the local volatility surface can be thought of as the market's expectation of the future value of volatility when the asset price is S at time t. Furthermore, the evolution of volatility along its path is condensed into a single function that is deterministic in S_t and t. For a give market level $K = S_t$ at $T = t$, the instantaneous volatility defined by (1.10Local Volatilityequation.1.3.10) is

$$\sigma(S_t, t) = \sigma_{S_t, t}(S_t, t).$$

We can then plug this expression for σ into equation (1.1Implied Volatilityequation.1.2.1) and retrieve the new dynamics of S_t

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma_{S_t, t}(S_t, t)dW_t. \tag{1.12}$$

Whereas (1.1Implied Volatilityequation.1.2.1) includes all possible models of volatility, such as stochastic, the diffusion parameter in (1.12Local Volatilityequation.1.3.12) is a one-factor model limited with a deterministic volatility function. In the next section we show how Dupire's LV is derived from the Fokker-Planck PDE.

1.4   PDE approach to Local Volatility

In this section we will derive the LV function from its PDE counterpart. We will follow a similar methodology used in the literature [47] and [23]. Our main approach relies heavily on recovering the call price from the risk-neutral density, Φ. Following Fengler's dual PDE approach, we assume the instantaneous volatility to be a deterministic function of S_t and T. By considering a one-factor diffusion process there exists a dual relation to the BSM PDE.

Under the risk-neutral measure $\mathbb{Q}$, the dynamics of S_t are governed by:

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \underbrace{\sigma(S_t, t)}_{dependant} dW_t,$$

where $\mu(\cdot)$ is the drift of S_t defined as the risk-free rate r minus the dividend yield q. Indeed the main difference here compared to the dynamics of equation (1.1Implied Volatilityequation.1.2.1) is that $\sigma(S_t, t)$ is now a dependent function in both spot and time. Using a probabilistic approach, we can obtain the BSM call option formula (1.7Implied Volatilityequation.1.2.7) by discounting the expected payoff under the risk neutral measure $\mathbb{Q}$. That is,

$$C_t(K, T) = e^{-r\tau} E_{\mathbb{Q}}[\max\{S_T - K, 0\}|S_t, t] \tag{1.13}$$

$$= e^{-r\tau} \int_0^\infty \max\{S_T - K, 0\}\Phi(K, T|S_t, t)dK. \tag{1.14}$$

Breeden and Litzenberger [7] first proposed that differentiating the call price with respect to the exercise price K twice yields the undiscounted probability distribution function known as the risk-neutral density (RND).

$$\Phi(K, T|S_t, t) = e^{r\tau}\frac{\partial^2 C_t(K, T)}{\partial K^2} \tag{1.15}$$

The state price density, or risk-neutral density function is the market's view of the future distribution of the call price C_t. The concept of a *transition* density is based on Arrow-Debreu Securities Debreu [11], (ADS). The prices of ADS are defined by the state-price density, which gives one dollar if the final state is in the interval $[x, x + dx]$ when starting from any point x and zero otherwise. The RND uniquely ensures an equivalent martingale measure under which all discounted asset prices are martingales. Another way of understanding the term state price density is by the following formula:

$$P(S_T \in [K_1, K_2]|S_t) = \int_{K_1}^{K_2} \Phi(K, T|S_t, t)dK. \tag{1.16}$$

Equation (1.16PDE approach to Local Volatilityequation.1.4.16) gives the probability that the stock is in the interval $[K_1, K_2]$ at time T, given the stock is already at level S_t in t. At first glance we can try to estimate the derivatives of Φ by a finite-difference scheme. However, this yields poor results since we do not have a continuum of option prices for every strike price, and the resulting RND is distorted as seen in Figure 1.7 (a). Also since we have to estimate the derivative twice this accentuates any errors produced from the first differentiation. For a detailed review on the various methods of estimating the RND we refer the reader to Grith et al. [25] and Chapter 11 of Jondeau et al. [29].

We may exploit the intrinsic relationship between implied volatility and the RND, following the approach of Benko [4]. An estimate of $\Phi(K, T|S_t, t)$ can be calculated as:

$$\hat{\Phi}(K, \tau) = F\sqrt{\tau}\Phi(d1)\left\{\frac{1}{K^2\beta_0\tau} + \frac{2d_1}{K\beta_0\sqrt{\tau}}\beta_1 + \frac{d_1d_2}{\beta_0}(\beta_1)^2 + 2\beta_2\right\}, \tag{1.17}$$

where β_0 is the interpolates IVS, and β_1, β_2 are the respective smoothed IV derivatives.

More on smoothing techniques can be found in Chapter 2. Two graphs of the RND are plotted in Figure 1.7.

(a) RND Numerical Differentiation   (b) RND Nonparamatric estimation
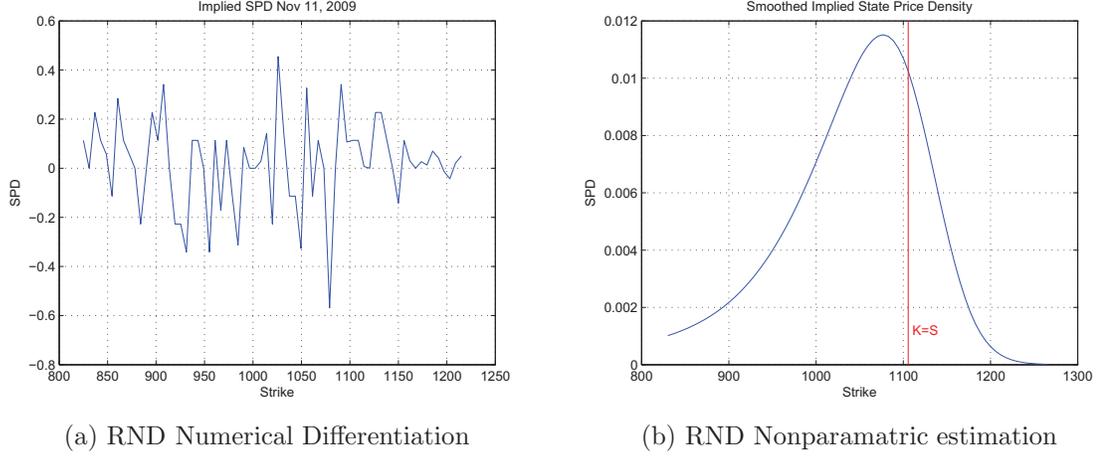
Figure 1.7: Left Panel: RND estimate via numerical differentiation yielding negative probabilities. Right Panel: RND via local polynomial regression with localized bandwidths.

The RND satisfies the BSM PDE

$$0 = \frac{\partial C}{\partial t} + (r - q)S_t \frac{\partial C}{\partial S_t} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S_t^2} - rC, \tag{1.18}$$

with terminal condition

$$\Phi(K, T | S_T, T) = \delta_K(S_T) = \frac{\partial^2}{\partial K^2}(S_T - K)^+,$$

where $\delta(\cdot)$ is the Dirac delta function. The transition density $\Phi(K, T | S_t, t)$ satisfies the *forward Kolmogorov* or *Fokker-Planck* PDE:

$$\frac{\partial \Phi(K^*, T | S_t, t)}{\partial T} = \frac{1}{2} \frac{\partial^2}{\partial (K^*)^2} \{\sigma^2(K^*, T)(K^*)^2 \Phi(K^*, T | S_t, t)\} \tag{1.19}$$
$$- \frac{\partial}{\partial K^*} \{\mu(S_t, t)K^* \Phi(K^*, T | S_t, t)\},$$

22

where S_t and t are assumed to be fixed. Equation (1.19PDE approach to Local Volatilityequation.1.4.19) spans across all maturities T and strike prices K^* with initial condition

$$\Phi(K^*, t|S_t, t) = \delta_S(K^*).$$

The use of the dummy variable K^* is needed for the integration below. The next step is to substitute (1.15PDE approach to Local Volatilityequation.1.4.15) in the above PDE. Using a simple chain rule from calculus, the first term becomes

$$
\begin{aligned}
\frac{\partial \Phi(K^*, T|S_t, t)}{\partial T} &= \frac{\partial}{\partial T}\left\{ e^{r\tau}\frac{\partial C_t(K^*, T)}{\partial K^2} \right\} \\
&= re^{r\tau}\frac{\partial^2 C_t(K^*, T)}{\partial (K^*)^2} + e^{r\tau}\frac{\partial^2}{\partial (K^*)^2}\frac{\partial C_t(K^*, T)}{\partial T}.
\end{aligned}
\tag{1.20}
$$

The second term in (1.19PDE approach to Local Volatilityequation.1.4.19) can be expressed as:

$$
\frac{\partial}{\partial K^*}\{\mu(S_t, t)K^*\Phi(K^*, T|S_t, t)\} = \mu(S_t, t)e^{r\tau}\frac{\partial}{\partial K^*}\left(K^*\frac{\partial C_t(K^*, T)}{\partial (K^*)^2} \right).
\tag{1.21}
$$

Equation (1.19PDE approach to Local Volatilityequation.1.4.19) can now be written as:

$$
\begin{aligned}
r\frac{\partial^2 C_t(K^*, T)}{\partial (K^*)^2} + \frac{\partial^2}{\partial (K^*)^2}\frac{\partial C_t(K^*, T)}{\partial T} &= \frac{1}{2}\left\{ \sigma^2(K^*, T)(K^*)^2\frac{\partial^2 C_{K^*, T}}{\partial (K^*)^2} \right\} \\
&\quad - \mu(S_t, t)\frac{\partial}{\partial K^*}\left(K^*\frac{\partial^2 C_t(K^*, T)}{\partial (K^*)^2} \right).
\end{aligned}
\tag{1.22}
$$

Integrating (1.22PDE approach to Local Volatilityequation.1.4.22) by parts twice from K to infinity gives us,

$$
\begin{aligned}
rC_t(K, T) + \frac{\partial C_t(K, T)}{\partial T} &- \frac{1}{2}K^2\sigma^2(K, T)\frac{\partial^2 C_{K, T}}{\partial K^2} \\
&+ (r - \delta)K\frac{\partial C_t(K, T)}{\partial T} - (r - \delta)C_t(K, T) = 0.
\end{aligned}
\tag{1.23}
$$

Rearranging (1.23PDE approach to Local Volatilityequation.1.4.23) we obtain the well known Dupire formula

$$\frac{\partial C_t(K,T)}{\partial T} = \frac{1}{2}K^2\sigma^2(K,T)\frac{\partial^2 C_t(K,T)}{\partial K^2} - (r-\delta)K\frac{\partial C_t(K,T)}{\partial K} - \delta C_t(K,T), \quad (1.24)$$

with initial condition $C(K,0) = (S_0 - K)^+$. When solved for $\sigma^2(K,T)$ gives equation (1.24PDE approach to Local Volatilityequation.1.4.24) the *Local Volatility Function*.

$$\sigma^2_{K,T}(S_t,t) = 2\frac{\frac{\partial C_t(K,T)}{\partial T} + \delta C_t(K,T) + (r-\delta)K\frac{\partial C_t(K,T)}{\partial K}}{K^2\frac{\partial^2 C_t(K,T)}{\partial K^2}}. \quad (1.25)$$

Hence a one-to-one mapping $\sigma(\cdot,\cdot) \leftrightarrow C(\cdot,\cdot)$ is given by the above equation along with the Kolgomorov equation (1.22PDE approach to Local Volatilityequation.1.4.22). The probabilistic approach to solving for the Dupire formula can be found in Section 3.3 of Fengler's book [17]. One of the main advantages of the Dupire equation is that it treats all call options as functions of strike K and maturity T which are observed at any given moment in time. Uniqueness is guaranteed by (1.22PDE approach to Local Volatilityequation.1.4.22) and hence can be used to calibrate a local volatility model for call (put) prices. A model is said to be well calibrated if it can reproduce market prices with very little error.

We have decided to exploit the unique relationship between the IVS and LVS by expressing σ_{LV} in terms of the IVS, $\hat{\sigma}$. Gatheral [23] and Wilmott [47] derive their own version of σ_{LV} by using the relationship between the *total implied variance* $w(S_t,K,T) = \sigma^2_{BS}(K,T)T$. For sake of clarity, we assume

- There exists a continuum of strikes K and maturities T for $\hat{\sigma}(K,T)$

- $\hat{\sigma}(K,T)$ is twice differentiable in K and once differentiable in T.

In terms of the moneyness metric κ, and time to maturity τ, we can express equa-

LVS SPX 11/23/2009
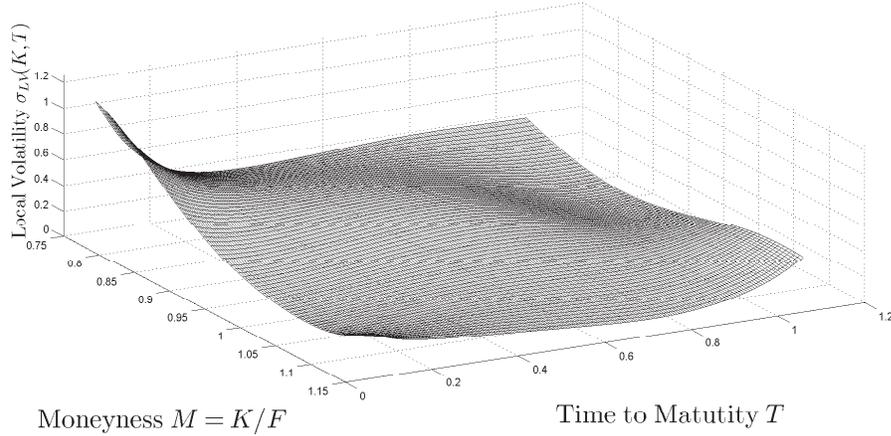
Figure 1.8: Local Volatility Surface on November 24, 2009 obtained from modified Dupire equation (1.25PDE approach to Local Volatilityequation.1.4.25) using $\hat{\sigma}$ obtained via local quadratic polynomial smoothing.

tion (1.25PDE approach to Local Volatilityequation.1.4.25) in terms of its BSM formula and its derivatives. Apply the chain rule on both the numerator and denominator of the Dupire formula and after some manipulation and cancelation of terms we get:

$$\sigma_{\kappa,\tau}^2(S_t, t) = \frac{\hat{\sigma}^2 + 2\hat{\sigma}\tau\frac{\partial\hat{\sigma}}{\partial\tau}}{1 + 2\kappa\sqrt{\tau}d_1\frac{\partial\hat{\sigma}}{\partial\kappa} + d_1 d_2(\kappa)^2\tau(\frac{\partial\hat{\sigma}}{\partial\kappa})^2 + \hat{\sigma}\tau(\kappa)^2\frac{\partial^2\hat{\sigma}}{\partial\kappa^2}}, \tag{1.26}$$

where

$$d_1 = (-\ln\kappa/\hat{\sigma}\sqrt{\tau}) + 0.5\sigma\sqrt{\tau}$$
$$d_2 = d_1 - \sigma\sqrt{\tau}.$$

For a collection of call prices $C_t(K, T)$ on the state space $\mathfrak{G}$ defined earlier, we can create a *Local Volatility Surface* much in the same way as we did for the implied volatility surface. Figure 1.8 displays the LVS for November 24, 2009. Due to the discreteness of the IV data, i.e., since we only observe σ_{IV} at a finite number of points (K, T), fitting a local volatility surface requires a strong from of interpolation. Indeed the problem of finding the LVS is "ill-posed", meaning that a small

change to the input (IVS) can lead to a large change in the output (LVS). There is a vast literature Crepey [10], Kahalé [32] on generating smooth LVSs. As is the case with the IVS there is no one method better than the other. Since our LV function is constructed from our IVS, we have used the coefficients generated from our local polynomial interpolation found in Chapter 2 to calculate the various derivatives of $\hat{\sigma}$. Other interpolation techniques include cubic spline interpolation of the call options with a Tikhonov penalization term [10], explicit and implicit finite differencing schemes [2], among others. A popular one being the Crank-Nicolson method. See [27] for a good understanding of approximating PDEs and their derivatives.

It is worth mentioning that local volatility can also be calculated using tree based algorithms. Derman and Kani [12] first proposed a binomial tree method based on forward induction. In essence this is just the undiscounted Arrow-Debreu prices for the discrete version of the RND. The same authors take it one step further and propose a trinomial tree in their paper [13] where the stock price can attain an extra level and does not change for $t + 1$ time. Another interesting property of the local volatility smile (fixed τ), again observed by Derman et al. [13] is that for a fixed time to maturity, the LV smile is approximately twice as steep as the IV smile. This behavior according to the authors is noticeable in equity markets and has been given the name the *two-times-IV-slope rule for local volatility*. A picture of this is displayed in Figure 1.9.

1.5   Data Analysis

Before moving forward with the various interpolation methods of the IVS, we present a detailed description of our data. We have used European call and put options from the S&P 500's options index; the SPX. The S&P500 Index is a capitalization-weighted
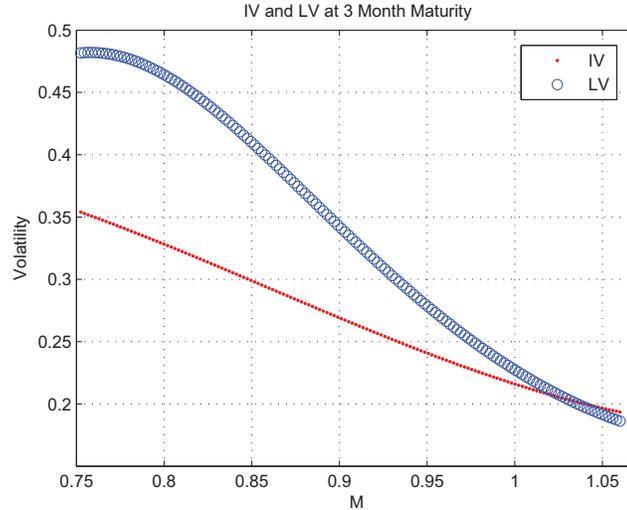
Figure 1.9: Interpolated IV slice(blue circle) Vs. LV slice(red square) at 3 month maturity. Local volatility slope is approximately twice as steep as the IV slope.

index of 500 stocks from a broad range of industries. More information on this index and others can be found on the Chicago Board of Options Exchange (CBOE) web-site[2]. The option prices used to back out the IVs are end-of-day (or last call) prices. Along with the option prices, the continuously compounded risk free interest rate and dividend yield is provided by WRDS' OptionMetrics[3] database.

We initially begin by screening our data for arbitrage violations. Any IV values that violate the boundary arbitrage conditions are removed. That is, all option prices less than or equal to the intrinsic value are removed, i.e, if $C_t \leq (S_t - K)^+$ and $P_t \leq (K - S_t)^+$. Similarly, we eliminate all options quoted less than $1/10^{th}$ of a dollar, those with an IV greater than 80%, and all options with a maturity less than 1 week because of their sensitivity to small errors in the data. We try to match the maturity of the option to that of the interest rate and use linear interpolation when interest rates for a specific maturity are not available. We have chosen a time to maturity (τ) grid between 0 and 365 days. From that we have selected to use options with

27

maturities [1,2,3,6,9,12] months. Our moneyness grid κ, ranges from 0.85 to 1.10, for an interval [0.85,0.90,0.95,1.00,1.05,1.10].

Let us recall that options are most liquid around the ATM region $\kappa \approx 1$ and less liquid the further away they deviate from that region. Only OTM options are used; $\kappa > 1$ for calls, and $\kappa < 1$ for puts. These are exactly the type of options that contain the most information about the implied volatility surface. As in Skiadopoulos et al. [44], ITM call and put options are not used because they have high deltas and therefore their IVs are very sensitive to the problem of non-synchronous data [26]. That is, often asset returns that are modeled as if synchronized from a timing perspective, but in reality they were produced with small delays or lags[4]. This effects prices derived from the BSM model and in turn their respective implied volatilities. Once we have fit our IVs using a nonparametric technique described in the next chapter, we group them in their respective ranges of moneyness and time to maturity. Thus, obtaining 36 time series of IV values per day, or six multivariate time series per implied volatility maturity denoted as $X_t(\kappa, \tau)$.

In our dimension reduction analysis we have decided to use daily log differences of smoothed implied volatilities, $\Delta X_t(\kappa, \tau) = lnX_t(\kappa, \tau) - lnX_{t-1}(\kappa, \tau)$. The reason behind this is that IVs often display a mean-reverting nature with the first autoregressive coefficient very close to unity. Hence, the time series of implied volatility returns are more often than not near stationary with a unitary root. Differencing once ensures autoregressive coefficients outside the unit circle and hence guaranteeing stationarity. After the procedures in the above paragraphs have been implemented our data is reduced from m observations to n observations. Table 1.1 displays summary statistics of our data.

---

[4]$http://www.bionicturtle.com/how-to/article/non-synchronous\_data/$

Table 1.1: Summary Statistics of IV log differences

Maturity Group	Mean	Stand. Dev	Min	Max
1	-0.0031	0.0717	-0.4894	0.2229
2	-0.0032	0.0670	-0.4855	0.2013
3	-0.0030	0.0625	-0.4766	0.1611
6	-0.0025	0.0620	-0.4106	0.2230
9	-0.0024	0.0643	-0.3951	0.2817
12	-0.0023	0.1086	-0.5644	0.5677

Note: Summary statistics of implied volatility returns computed as log differences as described above. The raw IVs are first fitted to a surface using a nonparametric technique as described in Chapter 2. The values for Mean and Stand. Dev are averaged across moneyness. Maturity group is measured in months.

We test the assumption of normality by using a standard Jarque-Bera test on IV returns for each of our 36 time series. That is for each point (κ_i, τ_j) in our series we run a chi-squared test statistic. The null hypothesis of normality is rejected for p-values less than 5%. Table 1.2 displays the results of this test for a sample of our data taken from August to December 2009. These results seem to be in line with those obtained by Skiadopoulos et al. [44] in their analysis of the IVS on the S&P500 index. Unlike the results obtained in Fengler and Hardle [18] on the German DAX index, the IV returns in our series almost always fail the test of normality.

The assumption of normality is further rejected by examining the Q-Q and cumulative distribution plots of the data in Figures 1.10 (a) and (b) respectively. Looking first at the Q-Q plot we see that the ends of the graph begin to break away from the straight line restriction imposed by Gaussian data. Next the cumulative distribution function is plotted along with that of a normal distribution. As in the previous figure we notice the tails are thicker compared to the Normal distribution.

Implied volatility is the only unknown parameter in the BSM equation. It is the

Table 1.2: Test for normality using Jarque-Bera test statistic

	Moneyness			
	0.85		1.10	
Maturity	χ^2	**p-value**	χ^2	**p-value**
1	2834.371	2.2e-16	613.5712	2.2e-16
2	4400.608	2.2e-16	890.6921	2.2e-16
3	5689.408	2.2e-16	1234.228	2.2e-16
6	1401.595	2.2e-16	460.6602	2.2e-16
9	635.43	2.2e-16	633.6861	2.2e-16
12	1416.304	2.2e-16	376.2169	2.2e-16

Note: Test for normality using a Jarque-Bera test statistic. The initial hypothesis is to assume the data is normally distributed. The test fails if the p-value is less than 5%.

(a) Q-Q Plot

(b) Cumulative Distribution Function
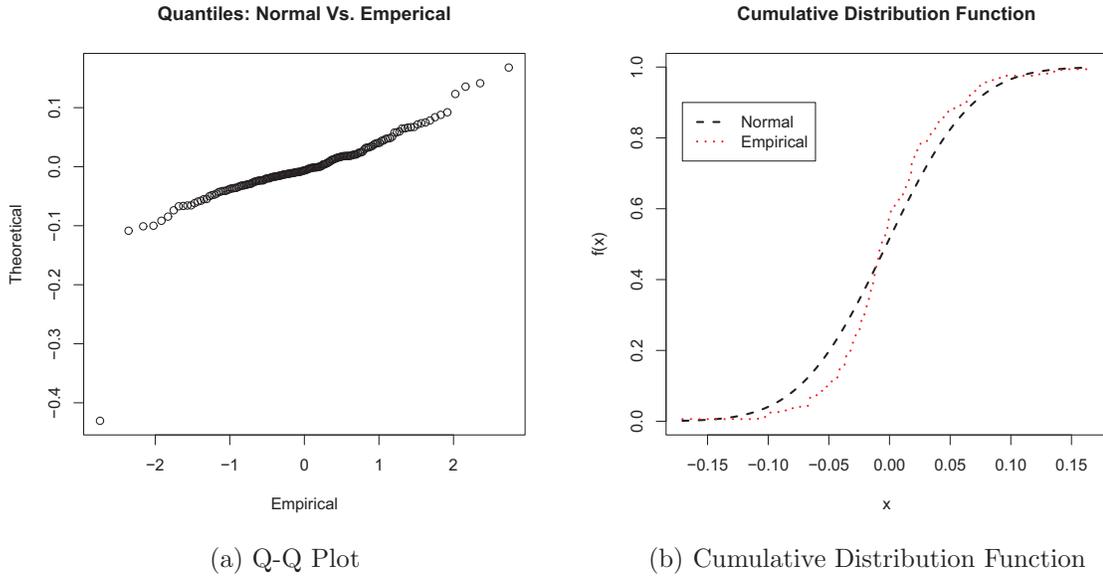
Figure 1.10: Q-Q Plot(left) and cumulative distribution(right) plot of the log differences of the IV at 1 month maturity and 0.85 level of moneyness.

solution σ_{IV}, which equates observed market option prices to those derived by the BS formula. As presented in this Chapter, IV is not constant as originally proposed. Furthermore, the convexity of IV allows us to express it as a map known as the *Implied Volatility Surface* (IVS). This representation allows us to exploit some interesting properties. The IVS displays a smile or smirk across the moneyness direction and

attains it maximum curvature around the ATM region. Across the time to maturity axes the curvature tends to flatten out for contracts maturing in the distant future. We introduced the theory of Local Volatility and its connection with the risk-neutral density function. The local volatility (LV) function is derived under the assumption that the diffusion parameter in our spot process is state-*dependant* as opposed to being a constant. This in turn can be used to efficiently price other derivative securities such as exotic options. The end of this chapter was devoted to describing our data and the various statistical tests that we ran. Our data consists of time-series vectors of IV log-returns for fixed levels of moneyness and time to maturity. As noted in the literature, IV follows an AR(1)process with an autoregressive coefficient close to 1 (hence the need to difference). Although the spot is assumed to be log-normally distributed, the IV time-series are highly non-normally distributed. In the next chapter we describe in detail the various tools and methods used to "calibrate" and construct an arbitrage-free IVS.

Chapter 2

Smoothing the Implied Volatility Surface

The problem with fitting the IVS is that we only observe IVs for a discrete set of moneyness and time to maturity. This is the string and pearl picture observed in Figure 1.2. In addition, our grid of observations changes as a function of the level of moneyness and time to maturity. Most of our data points seem to be concentrated around the ATM ($\kappa \approx 1$) region and become dispersed the further along we move along the time to maturity axis. Thus, we need to interpolate or smooth the data across both the moneyness and maturity dimension. There exists a vast literature on smoothing the IVS using parametric [42], semi-parametric [16], nonparametric and spline [48], [46] methods among others. In this thesis we have decided to use a nonparametric weighted regression to smooth our surface.

Another problem we are faced with once more is the presence of arbitrage. One of the main reasons why so much emphasis is put on interpolation of the IVS is because it does not "take" any specific functional form such as polynomial or nonparametric. Interpolation techniques often fall victim to arbitrage violations. The ones to

specifically avoid are negative vertical spreads, negative butterflies and negative calendar spreads. From what the literature suggests, there are two ways of creating an "arbitrage-free" IVS. In his paper, Kahalé [31] defends the need for the initial data to be arbitrage-free. He interpolates call prices by piecewise convex polynomials both in the time to maturity, τ dimension and strike, K dimension, and constructs an arbitrage-free surface in total implied variance $\sigma^2(\kappa, \tau)\tau$. This method is very useful when deriving the local volatility function and when accurately pricing exotic options. Fengler [15] argues that the input data need not be arbitrage-free, and instead proposes a B-spline method for recovering call prices on a finite grid resulting in an arbitrage free IVS, $\sigma^2(\kappa, \tau)$. For a well organized review of the IVS and its dynamics see Lee's paper [34] as well as Hull et al. [28]. In the following section, we will introduce the "natural" arbitrage bounds that arise from the construction of the IVS.

2.1   Natural Arbitrage Bounds of the IVS

The natural arbitrage bounds of the IVS are implicitly defined in the option price space or domain. Some basic definitions can be found below. For simplicity purposes we have only used call options in our definitions; however, they can be easily translated to include put options.

- **Vertical Call Spread** - Purchasing a number of call options and simultaneously selling an equal number of calls of the same class, underlying, expiration date but different strike price.

- **Calendar Spread** - Purchasing a number of call options and simultaneously selling an equal number of calls (usually ATM) of the same class, underlying, strike price but different expiration dates.

- **Butterfly Spread** - The butterfly spread is a risk neutral strategy. It involves buying a low strike ITM call, shorting two ATM calls and buying a high strike OTM call.

By definition, the call price $C(K,T)$ must be a monotonically decreasing, convex function in the K direction, and non-decreasing in the time to maturity direction:

$$-e^{-rT} \leq \frac{\partial C}{\partial K} \leq 0 \quad \text{and} \quad \frac{\partial^2 C}{\partial K^2} \geq 0. \tag{2.1}$$

The second inequality in (2.1Natural Arbitrage Bounds of the IVSequation.2.1.1) is the risk-neutral density function which was introduced in the local volatility section of Chapter 1. Moreover, for $t = 0$, $C(K,T)$ is bounded below by $\max\{S_0 - e^{-rT}K, 0\}$ and above by S_T, i.e.

$$\max\{S_0 - e^{-rT}K, 0\} \leq C(K,T) \leq S_T. \tag{2.2}$$

Let K_1 and K_2 be the strikes for two indexical call options. If $K_1 \leq K_2$, in order to avoid negative call spreads we must have:

$$C(K_1, T) \geq C(K_2, T) \quad \text{for } T \text{ fixed.} \tag{2.3}$$

Let T_1 and T_2 be the time to maturity for two identical call options. If $T_1 \geq T_2$, in order to avoid negative calendar spreads we must have:

$$C(K, T_1) \leq C(K, T_2) \quad \text{for } K \text{ fixed.} \tag{2.4}$$

In the following section we will present two methods to test for the presence of arbitrage; one using the RND, and the other using simple option-based strategies.

2.1.1 RND-based Tests

Arbitrage tests based on the RND involve checking whether the RND is a true density function. By definition of the density function, we require that Φ integrate to one and take on only positive values. Moreover, it should be able to reproduce the market call prices with very little error. The call prices obtained by integrating the RND should all be free of arbitrage, that is, the produced call prices should be decreasing monotonically in the strike direction. For the examples below, we have used option settlement data taken from the S&P500 on December 14, 2009 at 5 days to expiry. We have chosen such a short time to maturity for the specific purpose of finding arbitrage violations in call prices. Options with a very short time to maturity are more prone to errors. In our analysis, we have only used OTM call prices which guarantees us that calls with extremely high volatilities will be removed from our sample. Our approach is as follows:

- Obtain IVs from European call options

- Smooth the IV curve using a nonparametric technique

- Obtain fitted call prices using the smoothed IVs

Figures 2.1 (a) and (b) display the smoothed IVs and the call prices obtained using the sample IVs respectively. We must now check if the RND satisfies the above conditions. The area under the curve should integrate to 1 and should successfully recover the market prices. Next, we verify the monotonicity condition for each K_i.

$$\left.\frac{\partial C}{\partial K}\right|_{K=K_i} = -e^{-r\tau} \int_{K_1}^{\infty} \Phi(s)ds < 0. \tag{2.5}$$

(a) Smoothed IV smile   (b) Fitted Call prices
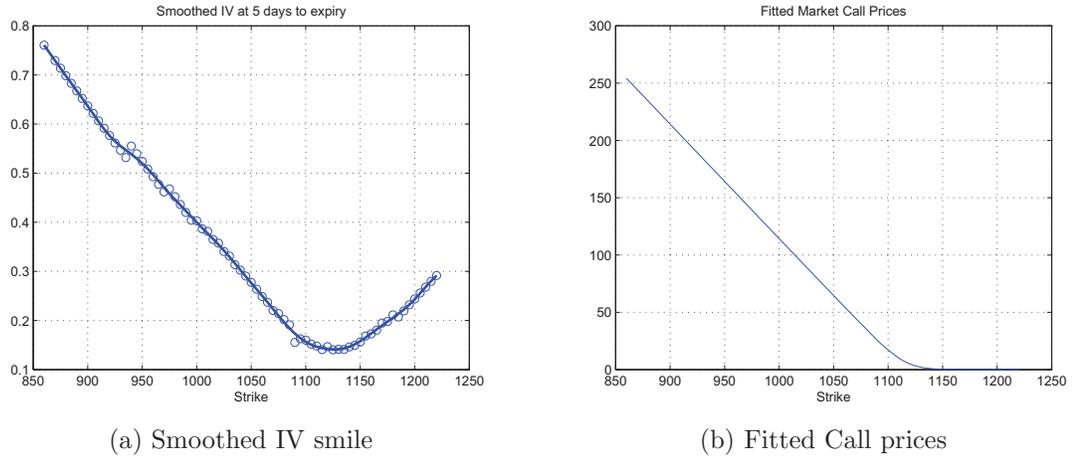
Figure 2.1: Left Panel: Smooth implied volatility smile for December 14, 2009 on the S&P500 with S0 = \$1114.11. Left axis denotes IV ticks and bottom axis denotes strike. Right Panel: Fitted market call prices using smoothed IVS. Smile has been smoothed using a local cubic polynomial with a Quartic Kernel function and localized bandwidth h_κ.

We also verify the convexity argument. For any two consecutive strikes, we must have the first derivative increase in strike:

$$\left.\frac{\partial C}{\partial K}\right|_{K=K_2} - \left.\frac{\partial C}{\partial K}\right|_{K=K_1} = e^{-r\tau} \int_{K_1}^{K_2} \Phi(s)ds > 0. \tag{2.6}$$

The results are summarized in the first four rows of Table 2.1. The RND integrates close to unity and reproduces the call prices fairly accurately. However, it does take on very small negative values close to zero. A plot of the RND for this data set is displayed in Figure 2.2.

2.1.2 Tests based on Option Strategies

Another way of testing for the presence of arbitrage is by using option-based strategies. We give a brief overview of the two strategies mentioned above, namely, vertical call bull spreads and butterfly spreads. We use a spacing of $dK = K_i - K_{i-1}$ for both tests. It is easy to show that at expiry the bull spread has a value in [0,1], that is,
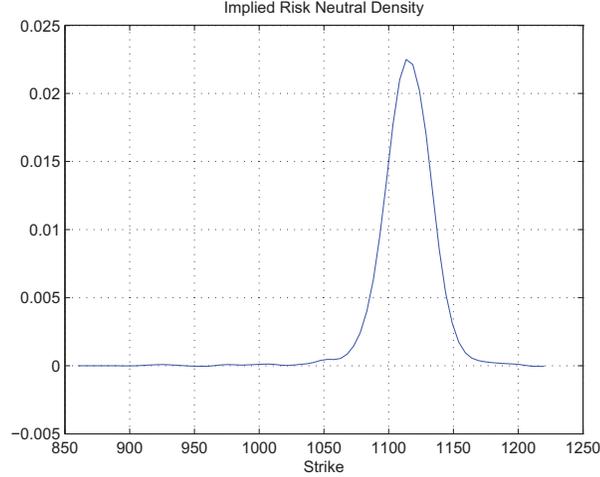
Figure 2.2: Risk Neutral Density recovered on December 14, 2009. Bottom axis denotes Strike price, K

$$VS_i = \frac{C_{i-1} - C_i}{dK} \in [0, e^{-r\tau}] \approx [0, 1] \quad \text{for } \tau \approx 0. \tag{2.7}$$

Similarly, a butterfly spread can be evaluated as

$$BS_i = \frac{C_{i-1} - 2C_i + C_{i+1}}{dK^2} \geq 0. \tag{2.8}$$

The results of these two tests appear in the bottom two rows of table 2.1. Taking into account the very short time to maturity, we clearly see that both tests fail the no-arbitrage conditions. 8% of the data produces calls with negative spreads and a whopping 21% fail the non-negativity of butterfly spreads. Although this might be seen as an extreme case, "fitting" an arbitrage free implied volatility curve that passes all of the above tests is not a simple task, albeit fitting an arbitrage-free surface is even more cumbersome. The problem is two fold. In one extreme, it can be seen that our initial data may contain errors, or that our smoothing technique caused the data to yield arbitrage induced values. For a detailed description on testing for arbitrage via the IV function see the paper [39].

Table 2.1: Summary of RND-based tests for Arbitrage

Statistic	Nonparametric Smoothing
Area under the curve(RND)	0.99
Call pricing errors(%)	22.22
Monotonicity errors(%)	12.34
Convexity errors(%)	9.12
Vertical Call Spread	8.45
Butterfly Spread	21.42

In his dissertation, Benko [4] uses a weighted least squares optimization to solve for an arbitrage-free IVS by combining the inherent relationship between the RND, the IV function, and its derivatives using equation (1.17PDE approach to Local Volatilityequation.1.4.17). This direct approach yields an IVS that respects the convexity argument, but fails to protect against call/put spreads and the general price bounds. For the sake of our analysis, we have decided to employ a non-constraint weighted least squares minimization to solve for the unknown volatility function $\hat{\sigma}$.

2.2   Nonparametric Methods

Let us consider the linear relationship between Y, the response variable and X, the predictor variable. The classical regression function take the form:

$$y_i = m(x_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.9}$$

$$m(x) = \mathbb{E}(Y|X = x). \tag{2.10}$$

In our context, these variables represent a form of moneyness measurement and time to maturity. We aim at estimating equation (2.9Nonparametric Methodsequation.2.2.9) in a nonparametric style. More precisely, our efforts are geared towards the proper estimation of $m(x)$ for a given data set $(x_i, y_i)_{i=1}^n$. Our use of the nonparametric

techniques lies in the assumption that the data around a local neighborhood of x contains some information of m at the point x. Hence, we seek an estimate $\hat{m}(x)$, by locally averaging the data around these points, such that:

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} w_{i,n}(x) y_i \tag{2.11}$$

where $w_{i,n}(x)_{i=1}^{n}$ denotes a sequence of weights. For $\epsilon > 0$ and x $\in$ X, more weight is given to values in the neighborhood of $(x - \epsilon, x + \epsilon)$ as opposed to the ones further apart. The choice of the weighting scheme depends on the individual and the problem at hand.

A kernel function will be used for our weights in equation (2.11Nonparametric Methodsequation.2.2.11). Kernel functions are positive, continuous, bounded, and symmetric which integrate to 1:

$$\int K(u)\, du = 1.$$

One can observe the obvious similarity to a marginal probability distribution function. We have decided to use the Gaussian kernel as our choice for $w_{i,n}$, which is given by:

$$K(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

The choice of the kernel function does not have a significant influence on the overall smoothing of the data. Our choice of kernel coincides with those chosen in the literature [9]. Given the multidimensional nature of our problem, we must interpolate both in the moneyness dimension and in the time to maturity dimension. Hence we obtain a two dimensional representation of the NW estimator via the product of kernel functions. Without loss of generality, assume that $K_1(u)$ and $K_2(u)$ are two independent Gaussian kernels, such that

$$K(u_1, u_2) = \prod_{i=1}^{2} K_{(i)}(u_i).$$

The success of the kernel depends heavily on the choice of bandwidth, h. For $h > 0$, the weighted kernel function is given by:

$$K_h(u) = \tfrac{1}{h}K(\tfrac{u}{h})$$

as well as satisfying the usual requirement:

$$\int K_h(u)\,du = 1.$$

2.2.1 The Nadaraya-Watson Estimator

Consider once more the general regression function

$$Y = m(X) + \epsilon,$$

where ϵ follows a white noise process, $\epsilon \sim WN(0, \sigma_\epsilon^2)$ s.t

$$\mathbb{E}(\epsilon|x) = 0 \quad \text{and} \quad \mathbb{E}(\epsilon^2|x) = \sigma^2(x).$$

Taking conditional expectations on both sides yields:

$$\mathbb{E}(Y|X = x) = E(m(X) + \epsilon|X = x) = m(x) + 0.$$

This in turn can be expressed in terms of the marginal pdf

$$m(x) = E(Y|X = x) = \tfrac{\int y f(x,y)\,dx}{f_x(x)}.$$

Let $(x_i, y_i)_{i=1}^n$ be our data set and replacing the integrals with sums, the NW estimator of $m(x)$ becomes:

$$\hat{m}(x) = n^{-1}\frac{\displaystyle\sum_{i=1}^n K_h(x - x_i)y_i}{\frac{1}{n}\sum_{i=1}^n K_h(x - x_i)}. \tag{2.12}$$

Rearranging equation (2.12The Nadaraya-Watson Estimatorequation.2.2.12) as

$$\hat{m}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{K_h(x - x_i)}{n^{-1} \sum_{j=1}^{n} K_h(x - x_j)} y_i = \frac{1}{n} \sum_{i=1}^{n} w_{i,n}(x) y_i,$$

reveals to us that (2.12The Nadaraya-Watson Estimatorequation.2.2.12) is nothing more than a local weighted average of the response variable y, with the kernel function acting as weights, such that

$$w_{i,n}(x) = \frac{K_h(x - x_i)}{n^{-1} \sum_{j=1}^{n} K_h(x - x_j)}.$$

Where the denominator is the normalizing constant which sums the weights equal to n. The conditional expectation and variance are given below:

$$\mathbb{E}[\hat{m}(x)] = \mathbb{E}\left[\frac{\sum_{i=1}^{n} K_h(x - x_i) y_i}{\sum_{i=1}^{n} K_h(x - x_i)} \right]$$

$$= \mathbb{E}\left[\frac{\sum_{i=1}^{n} (m(x_i) + \epsilon_i) K_h(x - x_i)}{\sum_{i=1}^{n} K_h(x - x_i)} \right]$$

$$= \frac{\sum_{i=1}^{n} m(x_i) K_h(x - x_i)}{\sum_{i=1}^{n} K_h(x - x_i)}.$$

Similarly the conditional variance is

$$\mathbf{V}[\hat{m}(x)] = \mathbb{E}[\hat{m}^2(x)] - [\mathbb{E}[\hat{m}(x)]]^2$$

$$= \mathbb{E}\left[\frac{\sum\limits_{i=1}^{n} y_i^2 K_h^2(x - x_i) + 2\sum\limits_{i=1}^{n} y_i y_j K_h(x - x_i) K_h(x - x_j)}{[\sum\limits_{i=1}^{n} K_h(x - x_i)]^2}\right] - [\mathbb{E}[\hat{m}(x)]]^2$$

$$= \mathbb{E}\left[\frac{\sum\limits_{i=1}^{n} K_h^2(x - x_i)(m(x_i)^2 + \sigma^2)}{[\sum\limits_{i=1}^{n} K_h(x - x_i)]^2}\right] - \frac{\sum\limits_{i=1}^{n} m(x_i)^2 K_h^2(x - x_i)}{[\sum\limits_{i=1}^{n} K_h(x - x_i)]^2}$$

$$= \frac{\sigma^2 \sum\limits_{i=1}^{n} K_h^2(x - x_i)}{[\sum\limits_{i=1}^{n} K_h(x - x_i)]^2},$$

where we have used the fact that $\mathbb{E}[\epsilon_i | X_i] = 0$ and $\mathbb{E}[K_h(x - x_i)\epsilon_i] = 0$.

The smoothing parameter h is the driving force behind nonparametric smoothing. Below are some examples of what happens to the estimator when h takes on different values. Note that in the limit as $h \to 0$ we get

$$\hat{m}(x_i) \to \frac{K_h(0)y_i}{K_h(0)} = y_i. \tag{2.13}$$

Equation (2.13The Nadaraya-Watson Estimatorequation.2.2.13) reveals that at worst, under smoothing just results in obtaining the response variable again. The other extreme, as $h \to \infty$.

$$\hat{m}(x_i) \to \frac{\sum_{i=1}^{n} K_h(0)y_i}{\sum_{i=1}^{n} K_h(0)} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}.$$

reveals that the bigger the parameter h is, the closer one gets to the sample mean by over smoothing.

2.2.2 Optimal Smoothing Parameter

As it is often the case when searching for an optimal value of a given parameter, we are faced with the choice between bias and variance. We define the smoothing window as $[x_0 - h(x_0), x_0 + h(x_0)]$ for the estimate of $\hat{m}(x)$, where the kernels are used as weights. The trade off between bias and variance can be expressed mathematically by searching for the value x that minimizes the mean square error (MSE). More precisely, for a random variable x, $\mu = E[x]$, and $\beta = E[x] - \mu$ the MSE is defined as:

$$\text{MSE}\{\hat{m}(x)\} = E[(\hat{m}(x) - m(x))^2] = E[(\hat{m}(x - \mu)^2] = \beta^2 + \sigma^2, \tag{2.14}$$

where the bias is given by $\hat{m}(x) = \mathbb{E}[\hat{m}(x)] - m(x)$. The Nadaraya-Watson estimator is biased in the sense that $\mathbb{E}[\hat{m}(x)] \neq \mathbb{E}[m(x)]$. Minimizing the distance of (2.14Optimal Smoothing Parameterequation.2.2.14) in the $\mathbb{L}^2$ sense can be achieved by penalty functions like Cross-Validation, General Cross-Validation, AIC and other forms of minimization techniques. In our case the choice of penalty function is motivated by the average squared error (ASE):

$$\text{ASE(h)} = \frac{1}{n} \sum_{i=1}^{n} \{\hat{m}(x_i) - m(x_i)\}^2 \widetilde{w}(x_i).$$

The ASE can be seen as a global measure to finding the optimal h by reducing the dimensionality of the equation. The ASE leads us to employ a "Leave-One-Out" Cross-Validation method described as:

$$\text{CV(h)} = \frac{1}{n} \sum_{i=1}^{n} [y_i - \hat{m}_{-j}(x_i)^2] \widetilde{w}(x_i),$$

where $\hat{m}_{-j}$ is the *leave-one-out* estimator. Essentially this follows a re-substitution procedure where the jth observation is left out. For a detailed review on the many ways of bandwidth selection we refer the reader to the work by Bowmman and Azzalini [6], as well as [41] and [8].

2.2.3 Local Polynomial Smoothing

We have decided to smooth the IVS using a local polynomial regression. The idea behind local polynomial smoothing is to try to fit a polynomial of order n to the regression equation. For this reason, the one dimensional Nadaraya-Watson estimator can be seen in the least squares sense defined by the minimizer

$$\hat{m}(x) = \underset{m \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - m)^2 K_h(x - x_i). \qquad (2.15)$$

Equation (2.15Local Polynomial Smoothingequation.2.2.15) defines a local linear polynomial regression. Solving (2.15Local Polynomial Smoothingequation.2.2.15) for the Normal equations yields (2.12The Nadaraya-Watson Estimatorequation.2.2.12) for m. Assuming that $m(x)$ is a continuous function up to order p, we can expand it via a Taylor approximation:

$$m(\zeta) \approx m(x) + m'(x)(x - \zeta) + ... + \frac{1}{p} m^{(p)}(x)(x - \zeta)^p$$

for ζ in the local neighborhood of x. Writing $m(x)$ as $\beta_0 + \beta_1 x$ and replacing it in (2.15Local Polynomial Smoothingequation.2.2.15) yields

$$\underset{m \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(x - x_i)\}^2 K_h(x - x_i). \qquad (2.16)$$

Again, using the kernel as weights, the local polynomial estimator of degree p of $\hat{m}(x)$ with coefficients $\beta_0, \beta_1, \ldots, \beta_p$ can be formulated by solving for the parameter β_0 that minimizes the following weighted sum of squares equation.

$$\hat{m}(x) = \underset{m \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(x - x_i) - ... - \beta_p(x - x_i)^p\}^2 K_h(x - x_i), \qquad (2.17)$$

where $\beta = (\beta_0, ..., \beta_p)^T$ is a vector of minimizing constants and $y = (y_1, ..., y_n)^T$ is a

vector of response variables. Equation (2.17Local Polynomial Smoothingequation.2.2.17)
can easily be solved by using matrix notation for matrices $\mathbf{X}$ and $\mathbf{W}$ defined as:

$$
\mathbf{X} = \begin{pmatrix}
1 & (x - x_1) & (x - x_1)^2 & \ldots & (x - x_1)^p \\
1 & (x - x_2) & (x - x_2)^2 & \ldots & (x - x_2)^p \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & (x - x_n) & (x - x_n)^2 & \ldots & (x - x_n)^p
\end{pmatrix}
$$

$$
\mathbf{W} = \begin{pmatrix}
K_h(x - x_1) & 0 & \ldots & 0 \\
0 & K_h(x - x_2) & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & K_h(x - x_n)
\end{pmatrix}
$$

Giving rise to the well-known least squares solution

$$
\hat{\beta}(x) = (\mathbf{X}^{\mathbf{T}}\mathbf{W})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{W}y \tag{2.18}
$$

where $\hat{\beta}_0$ gives the estimator $\hat{m}(x)$.

The choice of order p varies from problem to problem. We have chosen to use
$p = 2$ give by equation; however, odd degree polynomials tend to do better than even
degree ones since they better capture the peaks and valleys of the given function.
The Nadaraya-Watson estimator is a special case of equation (2.17Local Polynomial
Smoothingequation.2.2.17) with $p = 0$. Another popular choice is the local linear
polynomial of degree 1. Figure 2.3 displays call prices on November 24th 2009 which
have been fitted using a local linear estimator. More on local polynomial smoothing
can be found in Härdle's book [17].
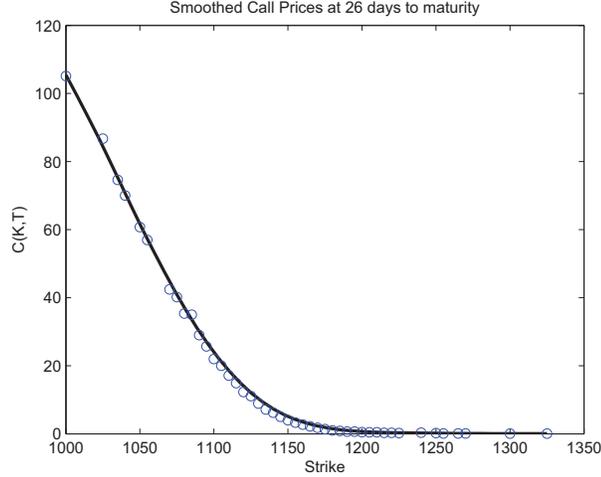
Figure 2.3: Smoothed Call Prices of November 24, 2009. Bottom axis denotes Strike price, K

It is interesting to note that as $h \longrightarrow \infty$ equation (2.16Local Polynomial Smoothingequation.2.2.1 collapses to the well known ordinary least squares regression. For a detailed proof of the conditional expectation and variance of the local linear estimator see the paper by Rupert and Wand [40]. After testing with various degrees we found that $p = 2$ best fits the data while capturing the most important properties. The two-dimensional local quadratic estimator of $m(x)$, is the value of the local regression curve

$$\beta_0 - \beta_1(x_1 - x_{1i}) - \beta_2(x_2 - x_{2i}) - \beta_3(x - x_{1i})^2 - \beta_4(x - x_{1i})(x_2 - x_{2i}) \qquad (2.19)$$

at x_0. Letting $x_1 = \kappa$ and $x_2 = \tau$ an estimate $\hat{m}(x)$, can be obtained by the weighted sum of squares:

$$\underset{m \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1(\kappa - \kappa_i) - \beta_2(\tau - \tau_i) - \beta_3(\kappa - \kappa_i)^2 - \beta_4(\kappa - \kappa_i)(\tau - \tau_i)\}^2 K_{h_1 h_2}(\kappa - \kappa_i, \tau - \tau_i),$$

$$(2.20)$$

where y_i represent the observed implied volatilities. The solution to (2.20Local Polynomial Smoothingequation.2.2.20) is given by

$$\hat{m}(x) = \hat{\beta}_0(x). \qquad (2.21)$$

46
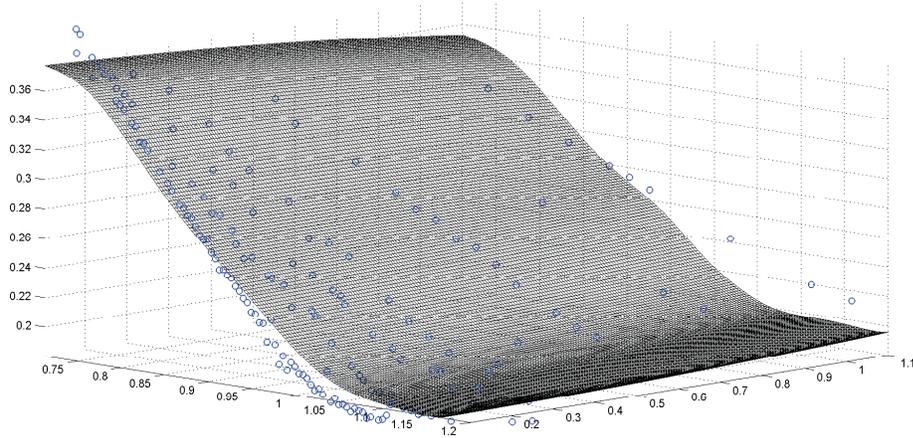
Figure 2.4: IVS for November 11, 2009 generated by Nadaraya-Watson estimate with a Gaussian Kernel. Left axis denotes moneyness, and right axis denoted time to maturity measured in years.

We have chosen to use a local weighted regression as opposed to the general NW estimator for a variety of reasons. Firstly, there is the problem of bias. More precisely we obtain better results when minimizing the MSE via local polynomial smoothing. For large κ and τ intervals, the IVS can be reasonably well fitted by piecewise polynomials. Secondly, as we move towards the wings of the IVS the points become more and more scattered as seen in Figure 1.2. In the case of NW smoothing we typically observe problems due to the one-sided neighborhoods at the boundaries (wings) as seen in Figure 2.4. The reason is that more or less the same points are used to estimate the curve near the boundary. Local polynomial regression; however, overcomes this by fitting a higher degree polynomial. If global bandwidths are to be used for both κ and τ, local polynomial regression outperforms the NW estimator even if the surface has been over smoothed. While choosing a bandwidth in the moneyness direction is often easily done, finding a bandwidth in the time to maturity is more cumbersome due to the lack of points, i.e, lack of long-term contracts. Fengler and Hardle [18] use a NW kernel estimate for smoothing the IVS surface by employing different bandwidths for T; h_{1T} is used for short term maturities and h_{2T} is used

for long term maturities. We have avoided this "splitting" of bandwidths via local quadratic polynomial smoothing. Polynomial smoothing is also less influenced by outliers in the data and therefore does not produce a "hump" in the graph as the NW estimator would.

Another useful property of polynomial regression is that we can explicitly define the derivatives of $\hat{m}(x)$. The derivatives of equation (2.20Local Polynomial Smoothingequation.2.2.20) are given by:

$$\hat{m}(x) = \hat{\beta}_0, \quad \hat{m}'(x) = \hat{\beta}_1, \quad \hat{m}''(x) = 2\hat{\beta}_2, \quad \ldots \quad , \hat{m}^p(x) = p!\hat{\beta}_p. \tag{2.22}$$

Using the above relations we can obtain the derivatives of $\hat{\sigma}$.

$$\beta_1 = \frac{\partial \hat{\sigma}}{\partial \kappa}(\kappa, \tau), \ \beta_2 = \frac{\partial \hat{\sigma}}{\partial \tau}(\kappa, \tau), \ \beta_3 = \frac{1}{2}\frac{\partial^2 \hat{\sigma}}{\partial \kappa^2}(\kappa, \tau), \ \beta_4 = \frac{\partial \hat{\sigma}}{\partial \kappa \partial \tau}(\kappa, \tau), \ \beta_5 = \frac{1}{2}\frac{\partial^2 \hat{\sigma}}{\partial \tau^2}(\kappa, \tau) \tag{2.23}$$

These were used to calculate the LVS of the Dupire [14] equation in Chapter 1. Specifically $\beta_1 = \frac{\partial \hat{\sigma}}{\partial \kappa}(\kappa, \tau)$, $\beta_2 = \frac{\partial \hat{\sigma}}{\partial \tau}(\kappa, \tau)$, and $\beta_3 = \frac{1}{2}\frac{\partial^2 \hat{\sigma}}{\partial \kappa^2}$ are needed to formulate equation (1.26PDE approach to Local Volatilityequation.1.4.26). Figure 2.5 displays the derivatives of $\hat{\sigma}$.

Backing out the implied volatility is only half the battle. Since options on the same underlying are not offered for a continuum of strike prices and time to maturity, a strong form of interpolation is then required to produce the IVS. As we saw in the beginning of this chapter, arbitrage violations are difficult to avoid even when the initial data (option prices) has been thoroughly screened. Furthermore, we saw that there are multiple ways to test for the presence of such violations via the RND and through the pricing bounds implied by the option prices. Once we have obtained the IVs by solving the inverse problem using the BSM equation, we subject them to the
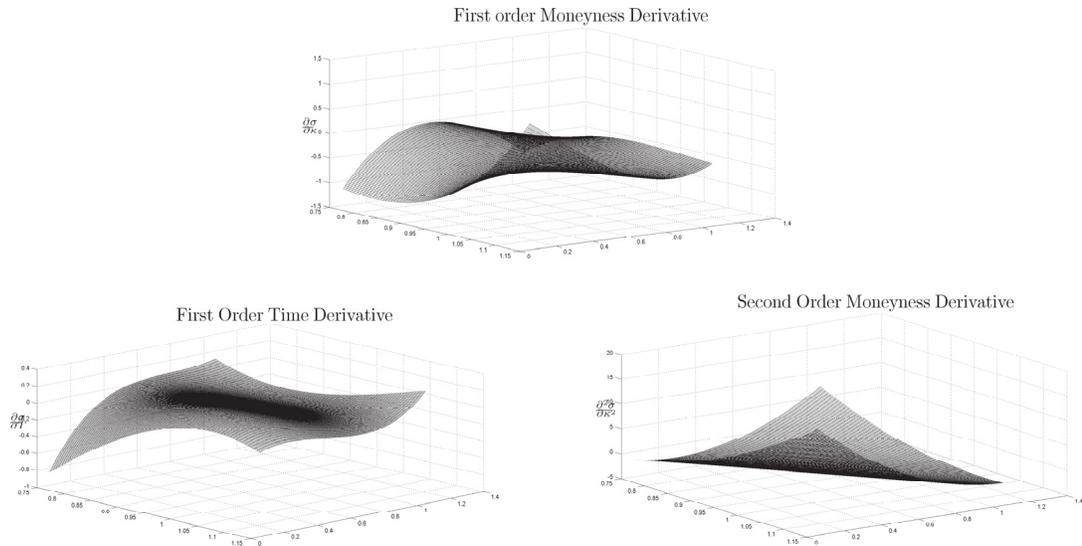
Figure 2.5: From left to right, the plots display derivatives of the IVS for SPX European options on 23/11/2009, first order moneyness derivative, first order time to maturity derivative, second order moneyness derivative using global bandwidths across moneyness and time to maturity

arbitrage tests described above before proceeding to forming the surface. There is no one perfect method, and almost all fail to produce a perfectly arbitrage-free surface. We then use a weighted local polynomial smoother which minimizes the most important types of arbitrage such as vertical and butterfly spreads and produce an overall smooth surface with minimal violations. As with any interpolation technique the bandwidth parameter is the key component to a successful smoothed surface. If over estimated it returns the sample mean, if under estimated than we obtain the response variable. An advantage of polynomial smoothing is that we can explicitly define the derivatives of our estimator. This proves to be extremely useful when defining the LV function.

Given the dimensionality of our data, a natural question to ask is whether all the variables play an equal role in describing the dynamics of the IVS. That is, can we breakdown the data is such a way as to retain the variables that have the highest

49

contribution to the surface? In the following chapter we will explore this question and present a series of results and interpretations.

Chapter 3

Dimension Reduction

One of the main problems in high-dimensional datasets is that, in most cases, not all variables are considered to be "important" for understanding the problem at hand. Dimension reduction methods aim at reducing the total number of variables of the given problem by extracting the variables that have the highest contribution to the variance of the data. We begin by giving a brief introduction to Principal Component Analysis, (PCA). PCA provides us with tools to reduce a high dimensional data set to a lower dimensional one while preserving the variables that describe the most of the variance. Thus revealing the hidden simplified structures. We will be applying PCA on slices of IV returns for fixed τ as in [44], and [9]. PCA takes its roots from simple linear algebra. Below is a short proof of PCA.

Let $\mathbf{X} = (X_1, \ldots, X_p)'$ be a random vector that has covariance matrix V with decreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$. For simplicity, assume that the data vector $\mathbf{X}$ is centered such that $\mathbb{E}[X] = 0$. Let $a \in \mathbb{R}$ be a vector of constraints. We are interested in finding combinations of $a_i \mathbf{X}_j$ such that the variances for each one is

as large as possible. Consider the linear combinations

$$Y_1 = \mathbf{a'_1 X} = a_{11}X_1 + a_{12}X_2 + \ldots a_{1p}X_p$$

$$Y_2 = \mathbf{a'_2 X} = a_{21}X_1 + a_{22}X_2 + \ldots a_{2p}X_p$$

$$\vdots = \vdots$$

$$Y_p = \mathbf{a'_p X} = a_{p1}X_1 + a_{p2}X_2 + \ldots a_{pp}X_p.$$

These linear combinations $Y_1, Y_2, \ldots, Y_p$ form the principal components of $\mathbf{X}$. Each of these Y_i are uncorrelated with one another and have maximum variance. The first principal component (PC) Y_1, is the linear combination with maximum variance $\text{Var}(Y_1) = a'_1 V a_1$. Since the variance can be increased indefinitely by increasing a, we restrict our vectors to having unit length. The idea is as follows:

1^{st}PC = Linear combination $a'_1 \mathbf{X}$ that maximizes $Var(Y_1) = a'_1 V a_2$ such that $a'_1 a_1 = 1$

2^{nd}PC = Linear combination $a'_2 \mathbf{X}$ that maximizes $Var(Y_2) = a'_2 V a_2$ such that $a'_2 a_2 = 1$

and $Cov(a'_1 \mathbf{X}, a'_2 \mathbf{X}) = 0.$

Continuing this way for the ith step we get,

i^{th}PC = Linear combination $a'_i \mathbf{X}$ that maximizes $Var(Y_i) = a'_i V a_i$ such that $a'_i a_i = 1$

and $Cov(a'_i \mathbf{X}, a'_j \mathbf{X}) = 0$ for $j < i$.

The following theorem summarizes the above with a detailed proof.

Theorem 3.1. Consider the data set $\mathbf{X} = (X_1, ..., X_p)'$. Define V the covariance matrix of $\mathbf{X}$ with eigenvalue-eigenvector pairs $(\lambda_1, \beta_1), \ldots, (\lambda_p, \beta_p)$ with the eigenvalues arranged in decreasing order. Then the ith *principal component* is given by the

projection

$$Y_i = \beta_i'\mathbf{X}, \quad i = 1, 2, \ldots, p,$$

such that,

$$Var(Y_i) = \beta_i'V\beta_i = \lambda_i,$$

$$Cov(Y_i, Y_j) = \beta_i'V\beta_j = 0, \quad i \neq j.$$

Proof. Proof 1 Using Lagrange multipliers

Let V be a covariance matrix of random variables $\mathbf{X} = (X_1, ..., X_p)$' such that V is symmetric and positive-semidefinite and assume for simplicity that $\mathbb{E}[\mathbf{X}] = 0$. Our problem is as follows; We are interested in finding a linear combination $Y = \alpha'\mathbf{X}$ for $\alpha \in \mathbb{R}$ such that the variance of Y is maximized, i.e. $\text{Var}(\alpha'\mathbf{X}) = \alpha'V\alpha$ over α, with $\|\alpha\| = 1$ our normalizing constraint.

Let V be a $p \times p$ matrix and $\lambda \in \mathbb{C}$ an eigenvalue of V. If there exists a vector constraint $\alpha_{p \times 1} \neq 0$ such that

$$V\alpha = \lambda\alpha, \tag{3.1}$$

then α is called an eigenvector of V corresponding to the eigenvalue λ. Equation (3.1Dimension Reductionequation.3.0.1) can be rewritten in terms of it's characteristic equation as

$$(V - \lambda I)\alpha = 0,$$

such that λ is a solution to $\det(V - \lambda I) = 0$. Below are some useful properties of the covariance matrix V.

- If V is symmetric then all eigenvalues $\lambda \in \mathbb{R}$ such that $\lambda_1 \geq \lambda_2... \geq .\lambda_p$.

53

- Trace$(V) = \lambda_1 + \ldots + \lambda_p$.

- $\det(V) = \lambda_1 \times \ldots \times \lambda_p$.

- $(\alpha')V\alpha = (\alpha')\lambda\alpha$ by equation (3.1Dimension Reductionequation.3.0.1) implies $\lambda = \frac{(\alpha')V\alpha}{\|\alpha\|^2}$.

Now let $g(x,y) = 1 - \alpha'\alpha$ be our constraint and consider the Lagrange multiplier λ such that

$$\frac{\partial(\alpha'V\alpha + \lambda(1 - \alpha'\alpha))}{\partial\alpha_i} = 0 \quad \forall \quad i = 1, \ldots, p.$$

Solving the above equation yields that λ is an eigenvalue of V and α its corresponding eigenvector such that $\lambda = \alpha'V\alpha$.

Let $\lambda_1 = \max(\alpha'V\alpha)$, i.e, we set λ_1 as having the maximum possible variance in our linear combination with α_1 as eigenvector and $\|\alpha_1\| = 1$ our usual constraint. Now consider the linear combination of $\alpha'X$ that maximizes $\alpha'V\alpha$ with $\|\alpha\| = 1$ **and** $\alpha_1'\alpha = 0$. We introduce the second lagrange multiplier η such that

$$\frac{\partial(\alpha'V\alpha + \lambda(I - \alpha'\alpha) + \eta\alpha_1'\alpha)}{\partial\alpha_i} = 0 \quad \forall \quad i = 1, \ldots, p.$$

Therefore, solving the above equation implies that λ is an eigenvalue of V with unit eigenvector α_2 with $\alpha_2 \perp \alpha_1$. Continuing the above procedure up to α_{p-1}, then by induction we obtain a sequence of eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p$ with optimal characterization:

$$\lambda_{k+1} = \max_{\alpha'\alpha_j = 0, \|\alpha\| = 1} \alpha'V\alpha$$

where, $\alpha_i'\mathbf{X}$ is called the ith principal component of $\mathbf{X}$. $\quad\square$

We now give an alternative proof to PCA via singular value decomposition.

Proposition 3.1. Consider the data set $\mathbf{X} = (X_1, ..., X_p)'$ and assume for simplicity that the data is centered, i.e., $\mathbb{E}[\mathbf{X}] = 0$. Define $V = \mathbb{E}[\mathbf{X}\mathbf{X}']$ the covariance matrix of $\mathbf{X}$ with eigenvalue-eigenvector pairs $(\lambda_1, \beta_1), \ldots, (\lambda_p, \beta_p)$ with the eigenvalues arranged in decreasing order. Then the ith *principal component* is given by

$$Y_i = \beta_i'\mathbf{X}, \quad i = 1, 2, \ldots, p,$$

such that,

$$Var(Y_i) = \beta_i'V\beta_i = \lambda_i,$$

$$Cov(Y_i, Y_j) = \beta_i'V\beta_j = 0, \quad i \neq j.$$

Proof. Let V be the covariance matrix of of a random vector $\mathbf{X} = (X_1, ..., X_p)'$ and assume for simplicity that $\mathbb{E}[\mathbf{X}] = 0$. Denote the covariance matrix of X by $V = \mathbb{E}[XX']$. We are interested in finding a linear combination

$$Y = \alpha'\mathbf{X}, \quad \alpha \in \mathbb{R}^p, \tag{3.2}$$

such that the variance of Y is as large as possible. The problem is not well defined since one can augment the variance of Y by increasing the value of α. The usual approach to this problem is to consider a normalizing constraint on α such that

$$\alpha'\alpha = 1. \tag{3.3}$$

Since V is symmetric and positive definite and $\mathbf{X}'\mathbf{X}$ is of full rank p we can decompose V by the singular value decomposition theorem. V can then be written as

$$V = \sum_{j=1}^{p} \lambda_j \beta_j \beta_j' = \beta\Lambda\beta', \quad \text{(in matrix notation)} \tag{3.4}$$

where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_p)$ is a $p \times p$ matrix of eigenvalues and $\beta = (\beta_1, ...\beta_p)$ is an orthogonal $p \times p$ matrix where each column is an eigenvector corresponding to an eigenvalue in Λ. Λ can be represented as

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}.$$

Hence, by equation (3.4Dimension Reductionequation.3.0.4) V can be expressed in terms of its eigenvalues and eigenvectors. This leads us to being able to find the principal components without needing to calculate the covariance matrix.

Once more, since V is symmetric we can arrange the eigenvalues in decreasing order, $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p$. The eigenvectors of V form a basis of $\mathbb{R}^p$, and so we can express the vector α as

$$\alpha = a_1 \beta_1 + \dots + a_p \beta_p, \tag{3.5}$$

for some $a = (a_1, \dots, a_p)' \in \mathbb{R}^p$. Constraint (3.3Dimension Reductionequation.3.0.3) is satisfied since α is a linear combination of β and $\alpha \alpha' = 1$ holds. From the orthogonality of β it follows that $\beta' V \beta = \Lambda$, and the variance of $\alpha' \mathbf{X}$ can be written

as

$$Var(\alpha' X) = \alpha' Cov(\mathbf{X})\alpha \quad (Matrix\ property)$$

$$= \alpha' V \alpha$$

$$= a'(\beta V \beta a) \quad \text{by} \quad (3.4 Dimension Reduction equation.3.0.4)$$

$$= a' \Lambda a$$

$$= \sum_{j=1}^{p} \lambda_j a_j^2$$

$$\leq \lambda_1 \sum_{j=1}^{p} a_j^2$$

$$= \lambda_1 \quad (\text{since} \quad a'a = 1).$$

The above implies that the variance of any linear combination of $\mathbf{X}$ cannot exceed its largest characteristic root λ_1. Setting $\alpha = \beta_1$,

$$Var(\beta_1 \mathbf{X}) = \beta_1' V \beta_1 = \lambda_1.$$

Thus, we can define the first principal component as

$$Y_1 = \beta_1' \mathbf{X}.$$

The next step is to add a second constraint forcing α to be uncorrelated with Y_1. That is, we wish to maximize $Var(\alpha' \mathbf{X}) = \alpha' V \alpha$ with $\|\alpha\|=1$ such that

$$\alpha_1' \alpha = 0.$$

More precisely,

$$\text{Cov}(Y_1, \alpha) = \text{Cov}(\beta_1' \mathbf{X}, \alpha) = \alpha' V \beta_1 = 0.$$

By induction, the above is continued at the kth step such that we wish to maximize

$Var(\alpha'\mathbf{X}) = \alpha'V\alpha$ for $\alpha \in \mathbb{R}$ with $\|\alpha\|=1$ and $\alpha'\mathbf{X}$ is uncorrelated with all $k-1$ previous linear combinations of $\mathbf{X}$. Thus, the principal components of $\mathbf{X}$ are defined as

$$Y = \beta'\mathbf{X}. \tag{3.6}$$

PCA can therefore be viewed as a mapping from $\mathbb{R}^p$ to $\mathbb{R}$ given by the projections $\beta_i'\mathbf{X}$ whose importance is measured by the size of the marginal variance $\lambda_i=\mathrm{Var}(\beta_i'\mathbf{X})$ alongside these same projections. $\qquad\square$

Some interesting properties of PCA follow:

- $Cov(Y) = \beta'V\beta = \Lambda$, PCs are pairwise uncorrelated

- $\lambda_i = Var(\alpha_i'\mathbf{X}), \quad$ where the elements of α_i are called factor loadings.

- $\sigma_{total}^2 = Trace(V) = \sum_{i=1}^{p} Var(\mathbf{X}_i) = \lambda_1 + ... + \lambda_p.$

Our ultimate goal is to determine if indeed the first few principal components can account for most of the overall variance σ_{total}^2. Whether

$$\sum_{i=1}^{k} \lambda_i/Trace(V) \approx 1 \quad \text{for small k.} \tag{3.7}$$

Thus, the eigenvalues that have the highest contribution to the overall variance are retained. In essence, PCA transforms the data so that it can be expressed in terms of the patterns between them. Geometrically, these linear combinations represent the set of a new coordinate system obtained by rotating the original data with $Y_1, Y_2, \ldots, Y_p$ as the coordinate axes. This new axis represents the directions with maximum variability therefore providing a smaller and more parsimonious description of the data.

Although PCA does not require the data to follow a Gaussian distribution, inferences can be made from the sample components when the population data is Normally

Table 3.1: PCA of IVS for 2 maturity groups

Maturity Group	2 Month	3 Month
Sample Covariance	$S_1 = \begin{pmatrix} 0.3785 & 0.4031 \\ 0.4031 & 0.7256 \end{pmatrix}$	$S_2 = \begin{pmatrix} 0.2592 & 0.2573 \\ 0.2573 & 0.6533 \end{pmatrix}$
Sample mean	$\overline{x}_1 = (0.3908, 0.5644)\prime$	$\overline{x}_1 = (0.2582, 0.4553)\prime$
Matrix of Eigenvectors	$\hat{\Gamma}_1 = \begin{pmatrix} 0.5498 & -0.8353 \\ 0.8353 & 0.5498 \end{pmatrix}$	$\hat{\Gamma}_2 = \begin{pmatrix} 0.4427 & -0.8967 \\ 0.8967 & 0.4427 \end{pmatrix}$
Eigenvalues	$\hat{\lambda}_1 = (0.9909, 0.1131)\prime$	$\hat{\lambda}_1 = (0.7803, 0.1322)\prime$

distributed. Another important assumption that PCA makes is *linearity*. The data is assumed to lie in a vector space, and PCA projects this data onto a lower dimensional linear subspace. This assumption may not hold for data sets lying on complex curved surfaces and therefore, the use of PCA may no longer be justified. More information on non-linear dimension reduction techniques will be presented in Chapter 4.

3.1 A PCA study of the IVS

For our study of the IVS, we apply PCA to our multivariate data set $\sigma(\kappa_i, \tau_i)$. Once the data has been cleaned to remove outliers and arbitrage opportunities (see Chapter 2 on data analysis), we are then left with the smoothed IVs for each day t, denoted by X_t. We apply principal component analysis to the daily log differences of X_t

$$\Delta X_t(\kappa, \tau) = \ln X_t(\kappa, \tau) - \ln X_{t-1}(\kappa, \tau), \tag{3.8}$$

for a fixed time to maturity.

To get a better understanding of our analysis, Figure 3.1 displays the three most prominent eigenvectors for our entire range of moneyness, $\kappa = [0.85, 0.90, 0.95, 1.00, 1.05, 1.10]$. There is no definitive answer to how many components we should retain.
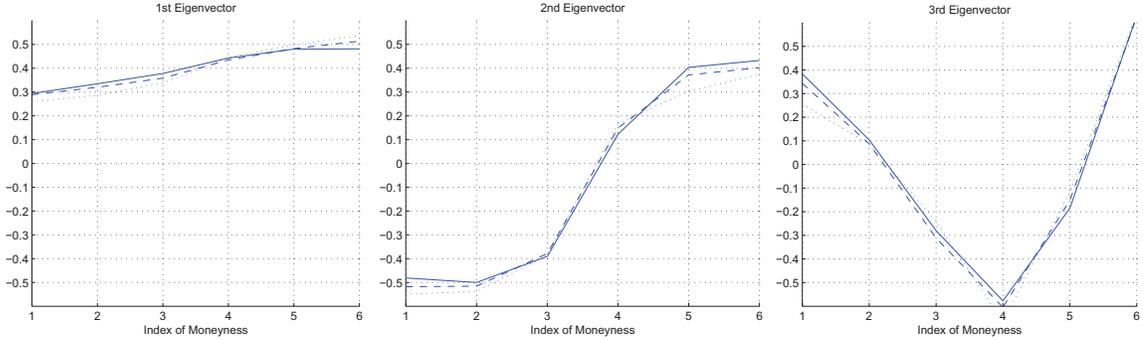
Figure 3.1: First, second, and third eigenvectors obtained from PCA for 1 month(solid), 2 month(dashed), and 3 month(dotted) maturity groups. Index of moneyness corresponds to the level of moneyness $\kappa \in [0.85, 0.90, 0.95, 1.00, 1.05, 1.10]$

Things to consider include the relative sizes of the eigenvalues, the amount of total variance explained, and other factors. We have followed the works of [9], [44], and [18]. A common technique in dimension reduction methods for identifying the number of PCs to keep is by graphing a scree plot. A scree plot is a plot of λ_i against i, the index number. To determine the number of components to retain we look for the bend in the graph. The "elbow" in the graph of Figure 3.2 (a) represents the magnitude (eigenvalue) that each component bears on the data. Looking at the scree plot we see that the 1st eigenvector is the most dominant, explaining 83% of the total variance as can be seen in Figure 3.2 (b).

Table 3.1 below displays our results from PCA applied to IVs with maturity 2 months and 3 months respectively. These results are found to be in accordance with the literature on this topic. The first eigenvector in Figure 3.1 has an almost flat slope and can be interpreted as the "shift" or "level" factor. The second component displays an almost Z-shape form and is called the "slope shock". The third eigenvector exhibits a twist or convexity change formation giving a heavier weight to at-the-money IVs. As can be seen in Figure 3.2 (a), the eigenvalues decrease in size and

similarly the variance explained (b) by each component decreases as well. The first three eigenvalues in the 1st maturity group explain 98.31% of the IVS variation. The percentage of variance explained is calculated by $\sum_{j=1}^{k} \lambda_{ij} / \sum_{j=1}^{p} \lambda_{ij}$, where the jth component refers to the maturity and the ith component refers to the level of moneyness. The same structure holds for the first three eigenvalues in all 6 maturity groups.

(a) Scree Plot   (b) Variance Explained
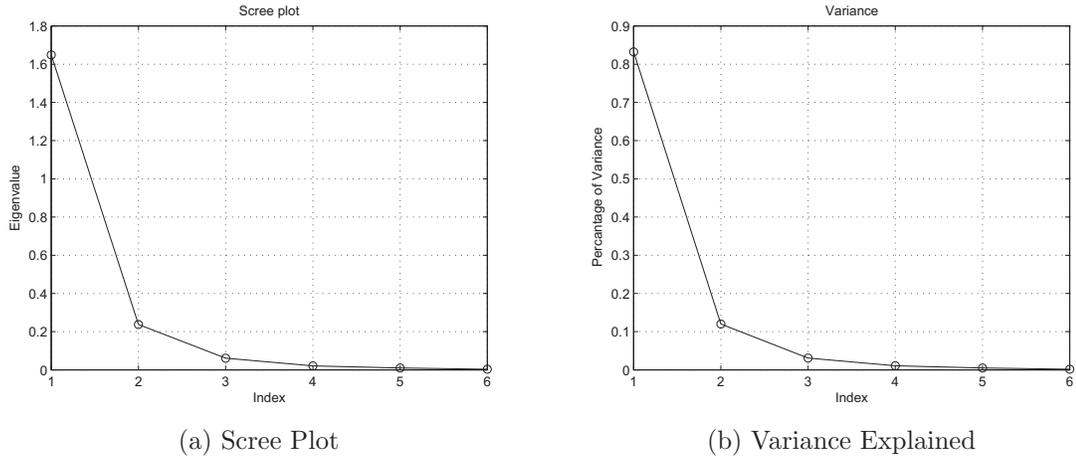
Figure 3.2: Left Panel: Scree Plot for IV differences with 30 days to maturity. The eigenvalues correspond to the variance attributed by each component and the Index 1-6 corresponds to the level of moneyness, $\kappa \in [0.85,\ 0.90,\ 0.95,\ 1.00,\ 1.05,\ 1.10]$ Right Panel: Variance explained of the kth component in group i defined as $\sum_{j=1}^{k} \lambda_{ij} / \sum_{j=1}^{p} \lambda_{ij}$.

We recall that the time series of IVs often display a mean-reverting property when modeled under an $AR(p)$ distribution with leading coefficient very close to one. Differencing our data corrects this issue and assures accuracy in our results. Several authors, for example [9, 44, 17], take it a step further and explore the relationships between the PCs and the underlying index. More precisely, they analyze the correlation between the first k most prominent PCs and the underlying asset. Figure 3.3 displays $x_1(t)$, the first principle component of the IVS for 40 days to maturity. It is observed to have a mean-reversion time of approximately one month. We have also

plotted the graph of the end of day log prices for the SPX index from January 1st, 2009 to December 31st, 2009. Even without any test statistics, the negative correlation is extremely evident. The authors use the evidence obtained from the autocorrelation function (ACF) and partial autocorrelation function (PACF) shown in Figure 3.4 and consider modeling $x_1(t)$ under an AR(1) process. Our analysis has shown that $x_1(t) \sim$AR(1) process with leading coefficient 0.8655. This near unity root confirms our need to difference the implied volatility returns.

(a) Projection of first principal component   (b) Log returns of the S&P500 Index
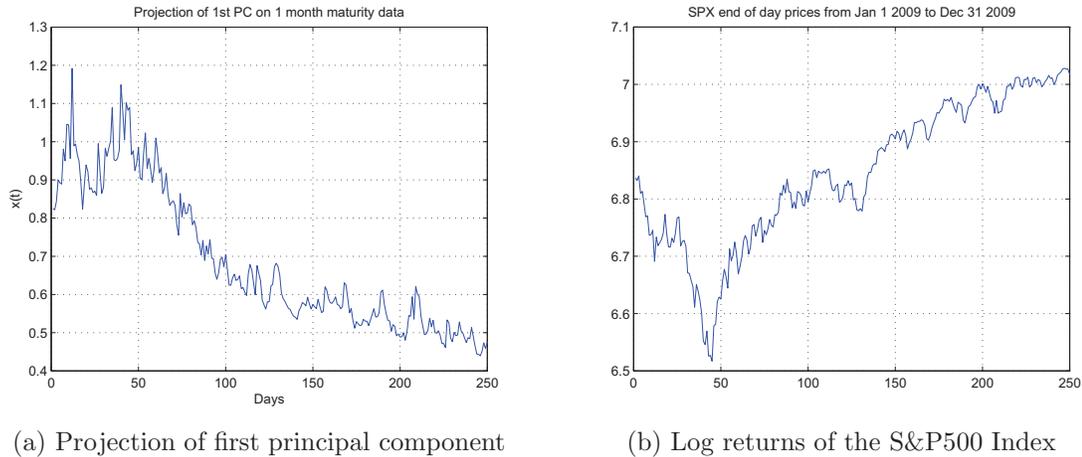
Figure 3.3: Left Panel: Projection of first PC on the IVS data with 40 days to maturity. Right Panel: Log returns of the S&P500 index.

As already observed, the implied volatility projected on the 1st PC moves in the opposite direction of the underlying. We compute the correlation coefficients between the differences of the 1st PC, $x_1(t) - x_1(t - 1)$ and the log difference of the underlying, $\ln$S$(t) - \ln$S$(t - 1)$ and find a correlation of -0.8121. These results are consistent with the leverage effect of the 1st eigenvector. The second PC; however, is positively correlated with the index. The second eigenvector in Figure 3.1 displays the slope effect which accounts for the general term structure observed in the IVS. In contrast to the first PC, the second PC displays a weaker correlation with the

(a) ACF of $x_1(t)$ (b) PACF of $x_1(t)$
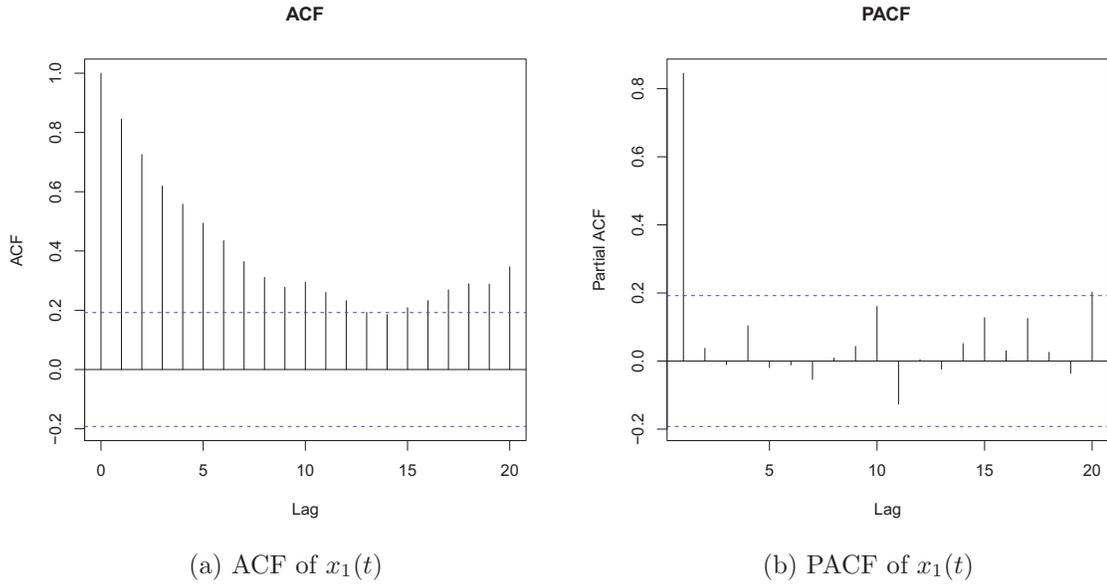
Figure 3.4: ACF (left) and PACF (right) of first principal component for 30 days to maturity. Blue horizontal lines indicate 95% confidence intervals.

underlying approximately equal to zero. Figure 3.5 displays the ACF and PACF of $\Delta x_2(t)$. There is presence of a mean-reverting process but with more jumps and spikes compared to the first PC. Running a model selection process reveals that $x_2(t)$ can be modeled by an ARMA(1,1) process. Finally, the 3rd PC has a "V-shaped" structure and can be interpreted as an ATM change in the convexity of the surface followed by a downward sloping term structure. Similar analysis reveals that $x_3(t)$ can also be modeled as an AR(1) process. These results are summarized in Table 3.2.

Table 3.2: Summary statistics of times series of principal components of S&P500 index options.

PC	Var. Expl(%)	Skewness	Kurtosis	Mean Reversion(days)	Corr. with Index
1	83	0.57	4.86	28	-0.8121
2	12	-0.10	4.10	15	0.0170
3	3	-0.63	3.43	24	-0.080

ACF

PACF

(a) ACF of $x_2(t)$

(b) PACF of $x_2(t)$
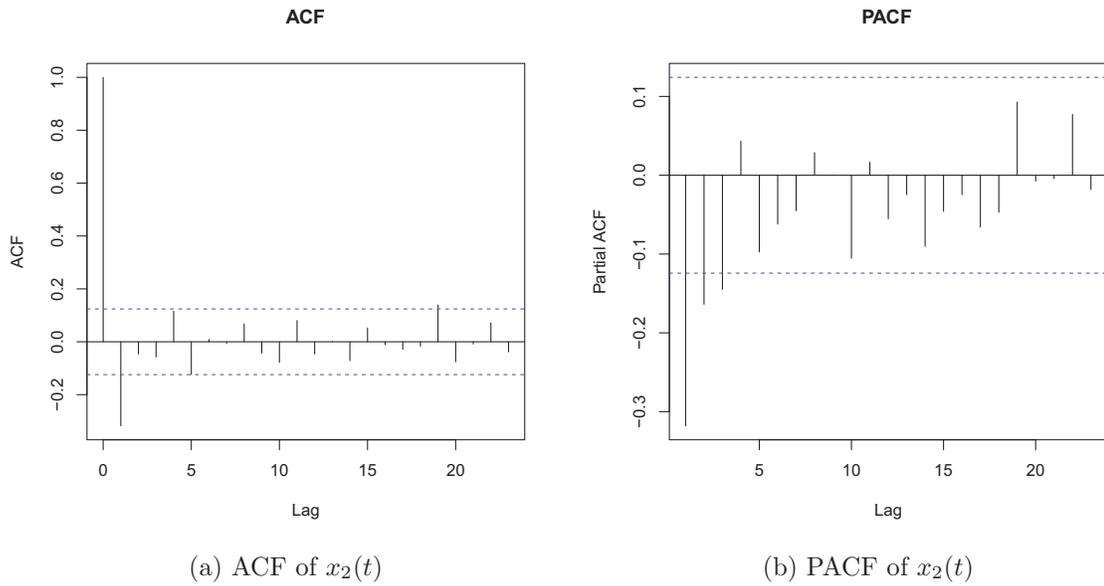
Figure 3.5: ACF (left) and PACF (right) of the differenced second principal component for 30 days to maturity. Blue horizontal lines indicate 95% confidence intervals.

ACF

ACF

(a) ACF of $x_3(t)$

(b) PACF of $x_3(t)$
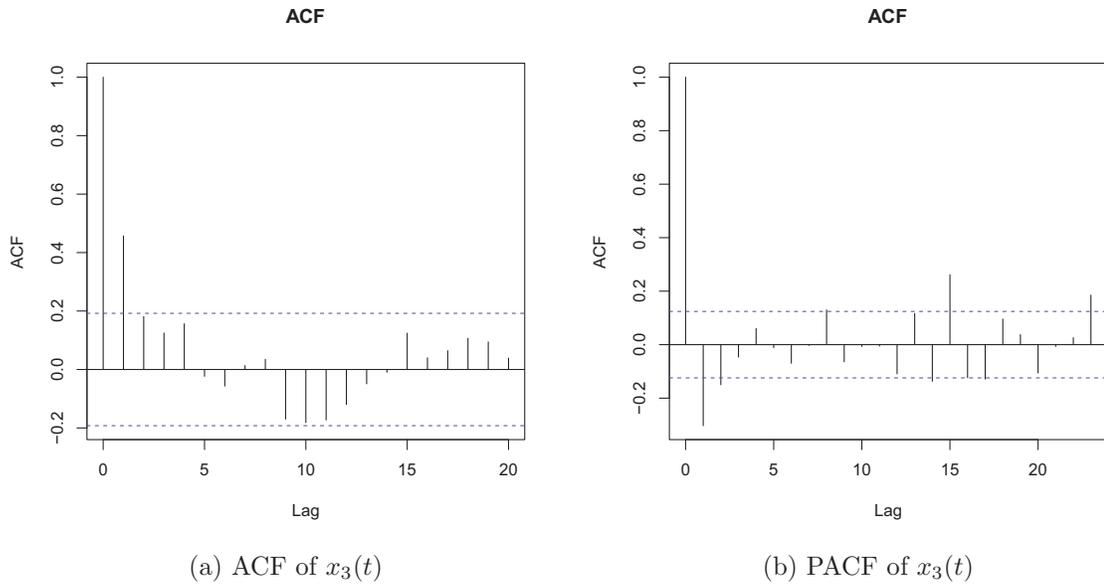
Figure 3.6: ACF (left) and PACF (right) of the differenced third principal component for 30 days to maturity. Blue horizontal lines indicate 95% confidence intervals.

3.2 Common Principal Components

There are a number of similarities arising from the covariance structure of IV returns. First, is that IV differences for short term maturities are more volatile and more dispersed along the grid than IVs for long term maturities. These similarities can be modeled in such a way where the eigenvectors are restricted to be *common* while the eigenvalues (variances) are allowed to change. Common principal component analysis (CPC) aims to capture this inherent similarity while holding no restriction on the variance. The following methodology follows Fengler and Hardle [18], and Flury [20], who introduced CPC and applied it to skull dimensions of Voles [21].

Consider the model associated with two covariance matrices Ψ_1 and Ψ_2 for two separate maturity groups indexed by 1 and 2:

$$\Psi_1 = \Gamma \Lambda_1 \Gamma^T \quad \text{and} \quad \Psi_2 = \Gamma \Lambda_2 \Gamma^T$$

where Γ is the matrix of eigenvectors and $\Lambda_i = \text{diag}(\lambda_{i1}, \lambda_{i2})$ the matrices of eigenvalues for maturity groups 1 and 2. The inspiration behind common modeling comes from the idea that the space spanned by the eigenvectors is identical across several maturity groups, whereas the variance of the individual components may vary. The covariance matrices satisfy the following commutative property

$$\Psi_1 \Psi_2 = \Psi_2 \Psi_1. \tag{3.9}$$

Naturally the question arises whether the principal components differ only due to sample variability. PCA only allows us to view slices of the IVS, say ATM IVs, therefore, only looking at one slice in time. However, CPC allows us to analyze several slices simultaneously for different maturity groups. This in turn, yields a joint eigenstructure across groups. Furthermore this technique reduces the dimensionality of the original problem by breaking down the data into a small number of factors that

are *common* across groups. This provides us with insight into the co-dependance between different maturities for a fixed level of moneyness.

The CPC analysis can be interpreted as follows. Consider covariance matrices $\Psi_1, \ldots, \Psi_k$, then the CPC hypothesis is formulated as:

$$H_{CPC}: \quad \Psi_i = \Gamma \Lambda_i \Gamma^T \qquad i = 1, \ldots, k \tag{3.10}$$

Let $\mathbf{X} = (X_{i1}, \ldots, X_{ip}) \in \mathbb{R}^p$, $i = 1, \ldots, k$, denote our smoothed log differenced IVs for k maturity groups and p grid (moneyness) points in the IVS. Ψ_i is our positive definite $p \times p$ population covariance matrix, $\Gamma = (\gamma_1, \ldots, \gamma_p)$ an orthogonal $p \times p$ matrix of eigenvectors, and $\Lambda_i = \text{diag}(\lambda_{i1}, \ldots \lambda_{ip})$ the matrix of eigenvalues. Denote by S, the sample covariance matrix of $\mathbf{X}$. Since $\mathbf{X}$ is assumed to follow a normal distribution such that

$$\mathbf{X} \sim N_p\,(\mu, \mathbf{\Psi}),$$

then the distribution of S follows a Wishart distribution with n degrees of freedom and parameter matrix Ψ/n or

$$S \sim W_p(n, \Psi_i/n).$$

A more detailed explanation on CPC and Wishart processes can be found in the book by Flury [20].

The number of parameters to be estimated for the CPC model are $p(p-1)/2$ for Γ plus kp for the eigenvalues Λ_i. The maximum likelihood estimates of Ψ_i are denoted by $\hat{\mathbf{\Psi}}_i = \hat{\Gamma} \hat{\Lambda}_i \hat{\Gamma}^T$. The sample principal components are computed via the usual projection $Y_i = \hat{\Gamma}^T \mathbf{X}_i$.

3.2.1 A CPC approach to the dynamics of the IVS

As in the case of PCA, we are reminded that we are working with a limited range of data. We observe IVs for a limited number of strike prices and an even thinner number of maturities. Hence, we need to interpolate across both the moneyness and time to maturity axes. Once a surface has been "fitted" to the data using interpolation techniques from Chapter 2, we obtain a time series of smoothed IVs for a range of moneyness κ_i, $i = 1, \ldots, 6$ and time to maturity τ_j, $j = 1, \ldots, 6$. We than take the log differences of the smoothed IVs, $\Delta X_t(\kappa, \tau)$ and group them into different maturity buckets. We obtain 6 maturities and 6 levels of moneyness for a total of 36 times series data.

The main advantage of CPC compared to ordinary PCA lies in the fact that it gives us a way to condense high-dimensional data into a small number of factor loadings that are *common* across different groups (maturities). In other words, we can examine the structure of the IVS for various levels of moneyness **and** different maturities. This in turn yields a smaller number of parameters to estimate compared to PCA. A motivation for CPC can be seen by re-examining the eigenvectors of Figure 3.1. There is little deviation of the overall shape for all three maturity groups, suggesting a joint structure. Figure 3.7 displays the three eigenvectors associated with the three largest eigenvalues under our CPC model for the 1 month and 3 month maturities. As is expected, all three eigenvectors retain the same structure as in the case for PCA. Table 3.3 displays the sample covariances S and estimated covariances Ψ obtained from CPC analysis for 1 month and 3 month maturity groups. Note that the commutative property of equation (3.9Common Principal Componentsequation.3.2.9) is satisfied up to a small numerical round off error. We have also plotted the scree plot for groups of 40, 60, and 90 days to maturity of the CPC model in Figure 3.8.

Table 3.3: CPC of IVS for 2 maturity groups

Maturity Group	2 Month	3 Month
Estimated Covariance	$\hat{\boldsymbol{\Psi}}_1 = \begin{pmatrix} 0.3435 & 0.3842 \\ 0.3842 & 0.7605 \end{pmatrix}$	$\hat{\boldsymbol{\Psi}}_2 = \begin{pmatrix} 0.3036 & 0.2813 \\ 0.2813 & 0.6089 \end{pmatrix}$
Matrix of Eigenvectors	$\hat{\Gamma}_{CPC} = \begin{pmatrix} -0.8594 & 0.5114 \\ 0.5114 & 0.8594 \end{pmatrix}$	
Eigenvalues	$\hat{\lambda}_1 = (0.1149 \,,\, 0.9891)'$	$\hat{\lambda}_2 = (0.1361 \,,\, 0.7763)'$

CPC analysis of first two maturity groups for a level of moneyness of 0.90 and 1.10. Sample covariance matrices are computed in table 1
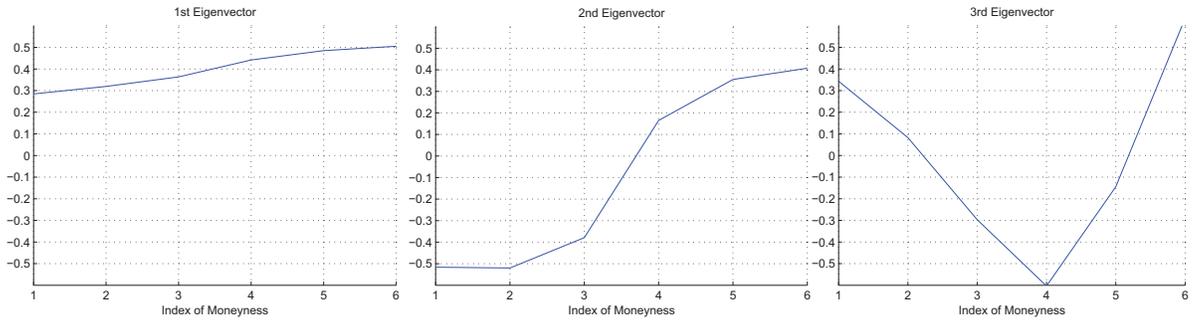
Figure 3.7: First, second, and third eigenvectors obtained from CPC for 1 month and 3 month maturity groups. Index of moneyness corresponds to the level of moneyness $\kappa \in [0.85, 0.90, 0.95, 1.00, 1.05, 1.10]$

Fengler and Hardle [18] expand the use of dimension reduction techniques by investigating other by-products of CPC. One such variant is the *proportional* model. The idea behind this model is that both covariance matrices are proportional up to a constant ρ such that all characteristic roots of $\Psi_1 \Psi_2^{-1}$ are the same and equal to ρ^{-1}. The hypotheses is presented as:

$$H_{PROP}: \quad \boldsymbol{\Psi}_i = \rho_i \boldsymbol{\Psi}_1 \qquad i = 2, \dots, k \qquad (3.11)$$

where $\rho_i > 0$ are the unknown constraints. Following the usual notation as above, we can express $\boldsymbol{\Psi}_i$ as $\boldsymbol{\Psi}_i = \beta \boldsymbol{\Lambda}_i \beta'$ with $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i1}, \dots, \lambda_{ip})$. This can be viewed as

the original CPC model with the following additional constraints

$$\lambda_{ij} = \rho_i \lambda_{1j}, \qquad i = 2, \ldots, k, \qquad j = 1, \ldots, p. \tag{3.12}$$

The number of parameters is $[p(p+1)/2]+(k-1)$. In relation to the IVS, this implies that variances of the common components between different groups are proportional to one another up to some positive constant say $\eta \geq 0$.

Regardless of the CPC model selected, common modeling is the right approach. Thus, we can split the the CPC analysis in two; one for short term maturities (1-5 months) and one for long term maturities (6-12 months). It is natural to ask whether we obtain the same eigenstructure when the surface has been smoothed using different bandwidths. The eigenstructure is preserved if local bandwidths are used. We use one pair of bandwidth (h_{Ms}, h_{Ts}) for the short term maturities and (h_{Ml}, h_{Tl}) for the long term maturities.

(a) Scree Plot   (b) Variance Explained
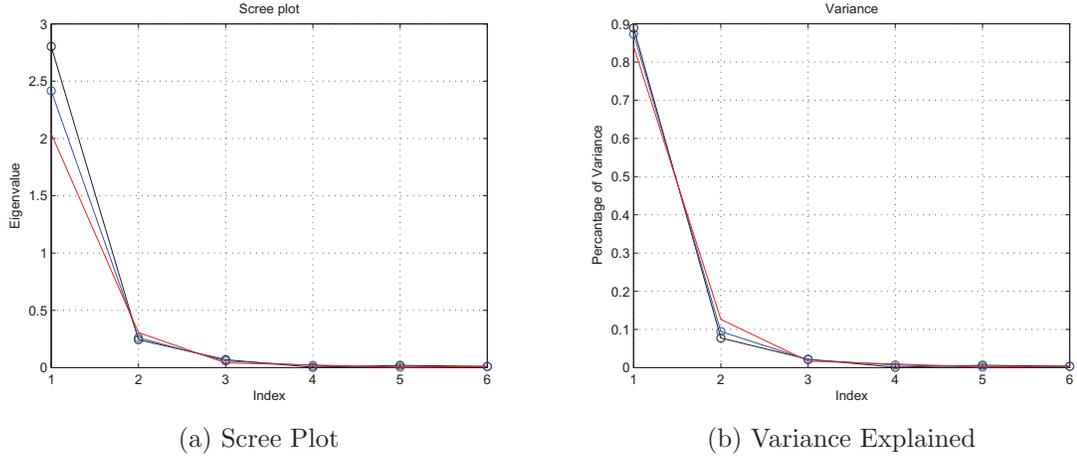
Figure 3.8: Left Panel: Scree Plot for IV differences with 40(black), 60(blue) and 90(red) days to maturity obtained via CPC. The eigenvalues correspond to the variance attributed by each component and the Index 1-6 corresponds to the level of moneyness, $\kappa \in [0.85, 0.90, 0.95, 1.00, 1.05, 1.10]$ Right Panel: Variance explained of the kth component in group i defined as $\sum_{j=1}^{k} \lambda_{ij} / \sum_{j=1}^{p} \lambda_{ij}$.

3.2.2 A reduced model

The main goal of dimension reduction methods such as PCA or CPC is to find a sequence of nested linear subspaces that best represents the *variability* of the data. In the same manner, we want to "choose" a volatility model reflective of the above property. Derman et al. [13] proposed a model of the IV smile for a fixed maturity and varying degree of moneyness known as the "sticky" model. For a fixed maturity τ_j and a range of moneyness $\{\kappa_1, \ldots, \kappa_n\}$ there exists a linear relationship between the log differenced IVs and a small number of common factors. These "*common*" factors can be taken to be the principal components obtained from the dimension reduction of the IVS. The three factors are the shift, slope and twist component.

For fixed maturity τ_j and range of moneyness $\{\kappa_1, \ldots, \kappa_n\}$ let $X_t(\kappa_i, \tau_j)$ denote the log differenced IVs as described in Chapter 2. The smile dynamics for a fixed maturity τ_j can be modeled as

$$\mathbf{X_t}(\tau_j) = \begin{pmatrix} X_t(\kappa_1, \tau_j) \\ X_t(\kappa_2, \tau_j) \\ \vdots \\ X_t(\kappa_n, \tau_j) \end{pmatrix} = \sum_{i=1}^{n} f_i(\tau_j) y_{ti}(\tau_j) + \epsilon_t(\tau_j), \qquad (3.13)$$

where y_t and ϵ_t are unobservable and assume to be i.i.d such that

$$\mathbb{E}[y_t(\tau_j)] = 0, \quad \mathbb{E}[\epsilon_t(\tau_j)] = 0, \quad \text{Cov}[y_t(\tau_i), \epsilon_t(\tau_j)] = 0.$$

Derman's "Sticky" model implies that the log differenced IVs can be modeled in such a way where they are only governed by a few driving factors. The model also displays a linear relationship between these movements. In our CPC framework, we can identify the variables $f_i(\tau_j) = \gamma_j$ as the common eigenvectors of $\hat{\Gamma}$ of the covariance matrix

$\hat{\Psi}$. The variance is naturally given by $\mathrm{Var}[y_j(\tau_i)] = \lambda_{ij}$. Hence, the entire surface dynamics can be described with three shock factors alone. If we were to use all the PCs in our decomposition of our factor model (3.13A reduced modelequation.3.2.13) then we would recover the original volatility surface.

We want to recover the unobserved signals or shocks, y_i from the data. We can generate these shocks by simulating from a multivariate Normal distribution. For a fixed τ, we set the variance equal to $\mathrm{diag}(\lambda_{11}, \lambda_{21}, \lambda_{31})$ for $n=3$ shocks and choose say $j = 1$ for the first maturity group. Using equation (3.13A reduced modelequation.3.2.13), our surface can be represented by these 3 principle components

$$\Delta X(\tau_1) = f_1(\tau_1)y_1(\tau_1) + f_2(\tau_1)y_2(\tau_1) + f_3(\tau_1)y_3(\tau_1). \tag{3.14}$$

Alexander [1] presents an interesting approach by applying PCA on fixed strike volatility deviations applied on a quadratic parametrization of the IVS as opposed to simple log-returns. Cont and da Foncesca [9] use an extension of (3.13A reduced modelequation.3.2.13) in their functional representation of the IVS.

$$X_t(\kappa_i, \tau_j) = X_0(\kappa_i, \tau_j) + \sum_{j=1}^{d} x_j f_j(\kappa, \tau), \tag{3.15}$$

where the PC's $x_j(t)$ are modeled as an AR(1) processes with interdependent white noise W_j, and $X_0(\kappa_i, \tau_j)$ is a constant surface. Another form of dimension reduction technique known as independent component analysis (ICA) is performed in the paper by Ané and Labidi [3] on actual surface data for a fixed level of moneyness and time to maturity rather than the log returns.

71

3.3 Review of dimension reduction methods

Dimension reduction techniques aim at reducing the dimension of the data while retaining the most important variables. In other words, it compresses the data into a smaller set by using the variables that have the highest contribution to the variance. Principal component analysis is one such method. It uses an eigenvalue-eigenvector decomposition and projects the data onto a lower dimensional linear subspace thus, revealing hidden structures within the data. These projections are called principal components (PCs). Our data consists of a daily times series of smoothed IVs. We apply PCA to the log-difference IVs, $\Delta X_t(\kappa, \tau) = \ln X_t(\kappa, \tau) - \ln X_{t-1}(\kappa, \tau)$, for a fixed time to maturity. We have kept 3 PCs as is common in the IV literature. These 3 PCs account for over 98% of the total variance in almost all maturity classes. Common principal component analysis (CPC) is similar to PCA but with one important difference, CPC finds a common eigenstructure across several *maturity* groups. CPC gives us the ability to analyze the covariance matrix of multiple maturity groups and find a common eigenvector while keeping the variance independent for all groups. Similar results are obtained using both reduction methods. The PCs can be interpreted as the slope, shift, and twist factors. The IVS can be entirely described by these 3 components given by the above mentioned factor model. These models alow traders or portfolio managers to hedge volatility risk, namely Vega-hedging. PCA and CPC rely heavily on the notion that the data is vector based, i.e., Euclidean. This assumption is often overlooked and in some cases may not even be true. We shall see in the next chapter that a certain class of volatility models inherit a Riemannian geometry where the data no longer lie in a vector space. A new form of dimension reduction is then required.

Chapter 4

Geometry of the Implied Volatility Surface

In this chapter we present a different approach to modeling the implied volatility surface. There is indeed a connection between Riemannian geometry and mathematical finance. We explore the concept of defining a mean on a manifold, and introduce a new form of dimension reduction known as principal geodesic analysis (PGA). We give an example by applying PGA to the sphere in $\mathbb{R}^3$. Most of the definitions and terminology used in differential and Riemannian geometry can be found in the Appendix. Our previous approach relied on estimating the variance in a Euclidean space when the problem is clearly non-Euclidean. We propose applying PGA on our data set; that is on actual manifold data. The main difference from traditional dimension reduction techniques, relies in the use of the inherent non-linear structure of the problem, and the use of an intrinsic mean.

4.1   Principal Geodesic Analysis

Like PCA, Principal Geodesic Analysis (PGA), introduced by Fletcher [19] is a dimension reduction method. While PCA is restricted to working in a Euclidean (linear)
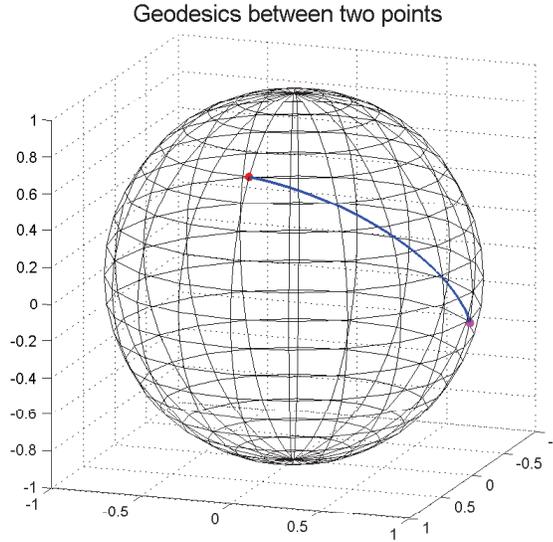
Figure 4.1: Geodesic curve between two points on the sphere. This is the unique minimizing geodesic such that it is the shortest distance between these two points.

vector space, PGA uses the inherent non-linear structure of the data to project onto a lower dimensional Riemannian manifold. Riemannian manifolds are used to enforce consistency in data and define more accurate metrics to work with. Definitions and terminology used in differential and Riemannian geometry can be found in the Appendix. The interest in manifold modeling arises from the non-linearity of the problem. Benefits include dimension reduction, accuracy in measurements, and consistency in model representation. The key ingredient in PGA is the use of geodesics. As stated in Definition 4.6 of the Appendix, geodesics can be thought of as the manifold generalization of a straight line, i.e., as the shortest distance between two points in the manifold.

For an illustrative example consider points p and q on the the 2-sphere manifold in $\mathbb{R}^3$ represented in Figure 4.1. The shortest distance between these two points is given by the geodesic in blue. This is the length-*minimizing* geodesic. In other words, it is the shortest path to travel while still remaining on the manifold. Recall, a manifold

is called Riemannian if it is endowed with a Riemannian metric g. Notions such as mean, variance, and distance can be defined on a manifold the same way they are defined in Euclidean space. The only drawback is that unlike their Euclidean counterpart, manifold statistics often do not admit a closed-from solution, and regularly require numerical methods to be implemented.

As in PCA, we seek to project the data onto a lower dimensional space. In PGA we project the manifold data onto the tangent space T_pM about the mean point of the manifold. PGA centers its operations about the mean point $\mu \in M$. We define two distinct notions of distance. The first being the classical Euclidean (extrinsic) distance and the second the Riemannian (intrinsic) distance. Given a set of points $x_1, x_2, \ldots, x_n \in \mathbb{R}^d$, the arithmetic mean or average, $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is defined as the point that minimizes the sum-of-squared Euclidean distances to the given points. In other words:

$$\operatorname*{argmin}_{x \in \mathbb{R}^d} = \min_{x \in \mathbb{R}^d} \sum_{i=1}^{n} \|x - x_i\|^2.$$

The norm refers to the Euclidean distance between points x and x_i which is the length of the line segment connecting them. In a sense, this can be thought of as a linear mean. The notion of distance on a manifold has a different meaning. Given that the manifold M may not necessarily form a vector space, this notion of mean is then rendered useless.

Since we are working on a manifold M, one way to define distance between points is to embed it in a Euclidean space. An embedding is a diffeomorphism of M onto its image such that the image of the embedding must be a submanifold N of M. The Nash embedding theorem states that every Riemannian manifold can be embedded into some Euclidean space while still preserving distance. Let $\Phi : M \to \mathbb{R}^d$ be such an embedding then we can define the notion of *extrinsic* mean of a collection of points

$x_1, x_2, \ldots, x_n \in M$ by

$$\mu_\phi = \operatorname*{argmin}_{x \in M} \sum_{i=1}^{n} \|\phi(x) - \phi(x_i)\|^2.$$

Once again though, we are linearizing the mean in the Euclidean sense. The data points are treated as if obtained from a vector space and the shortest distance projection is used to find the mean on the manifold. Since we want to retain the inherent structure of the manifold a natural candidate is the *intrinsic* mean. Assuming the same set of points as above, the intrinsic mean (IM), is the minimizer in M of the sum-of-squared Riemannian distance to each point

$$\mu = \operatorname*{argmin}_{q \in M} \sum_{i=1}^{n} d_R(x_i, q)^2 \tag{4.1}$$

where $d_R(x_i, q)$ denotes the Riemannian distance between the ith data point and the mean candidate q. This is the type of mean that we will be working with.

The geodesics of the sphere are given by the great circles passing through both poles. The sphere is a particularly good example to work with since the geodesics are defined explicitly. They are found by solving the Euler-Lagrange equations. Figure 4.2 displays one such geodesic passing through both poles at a particular point in time. Using the inherent geometrical properties of the sphere, we will show in the next section how the above minimization problem can be solved explicitly by utilizing the *Exp* and *Log* maps defined in the Appendix.

4.1.1   Intrinsic Mean on S^2

In this example we try to visualize the intrinsic mean for a set of points on the 2-sphere S^2 in $\mathbb{R}^3$. We consider the geodesics at the base point $p = (0, 0, 1)$, the north pole. Visually this translates to the meridians, or the great circles of the sphere. The sphere can be represented by the symmetric quotient space $M \simeq SO(3)/SO(2)$.
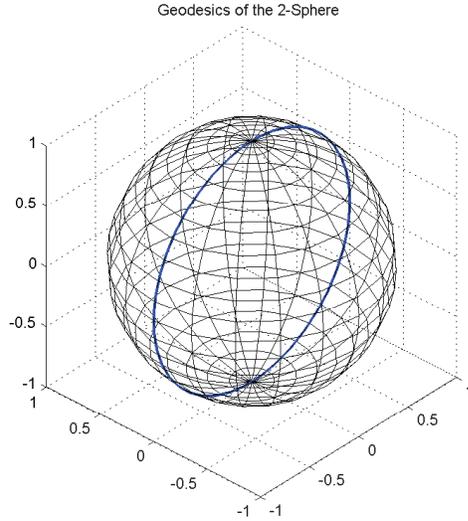
Figure 4.2: Geodesics on the sphere are given by the great circles (meridians) passing through both poles.

The geodesics of M are realized by the group action acting on the one-parameter subgroup, $SO(3)$. Consider a tangent vector $t = (t_1, t_2, 0)$ in T_pM in the xy plane, and recall, the exponential map at point p maps straight lines through (0,0) of T_pM to geodesics in M passing through p. Vectors in the tangent space are mapped back to M using this exponential map. The Riemannian exponential map of $SO(3)$ given by Rodriguez's formula is

$$\text{Exp}_p(t) = \left(t_1 \cdot \frac{sin\|t\|}{\|t\|}, t_2 \cdot \frac{sin\|t\|}{\|t\|}, cos\|t\| \right), \tag{4.2}$$

where $\|t\| = \sqrt{t_1^2 + t_2^2}$. In a similar fashion we can move from the manifold space to the tangent space with the Log map. For a point $x = (x_1, x_2, x_3) \in S^2$, the log map is given by

$$\text{Log}_p(x) = \left(x_1 \cdot \frac{\theta}{sin\theta}, x_2 \cdot \frac{\theta}{sin\theta} \right), \tag{4.3}$$

where, $\theta = \arccos(x_3)$ is the spherical distance from the base point p to the given point x. An illustration of this example is displayed in Figure 4.3 where we have generated

random points on a geodesic between points p and q and have calculated the IM (yellow) using equation (4.1Principal Geodesic Analysisequation.4.1.1).

Similar to Fletcher we have an algorithm for implementing PGA summarized in Algorithm 1.

Data: Points $x_1, \ldots, x_n$ in M.
Result: Approximating PGA
Calculate intrinsic mean;
while $\Delta q > \epsilon$ **do**
$\quad$ Set $q = x_1$;
$\quad \Delta q = \frac{1}{N} \sum_{i=1}^{N} Log_{q_j} x_i$;
$\quad q_j \rightarrow q_{j+1} = \text{Exp}_{q_j}(\Delta q)$;
end
Initialize PGA;
z=$\text{Log}_q(x_i)$;
S $= \frac{1}{N} \sum_{i=1}^{N} z_i z_i'$;
Extract principal directions and variances;
(β, λ);

Algorithm 1: PGA Algorithm

PCA is dependant on the *vector* space structure of the data and hence, cannot be used on manifold valued data. Utilizing the inherent geometry of manifolds leads us to generalize the notion of linear subspaces found in PCA to *geodesic submanifolds* used in PGA. The geodesic curve is the Riemannian analog of the first principal direction in PCA. We require the submanifolds to be *geodesic* for the following reasons. If N is a submanifold of M, then, geodesics of N are not necessarily geodesics of M. A submanifold N of M is said to be geodesic at $x \in N$ if all geodesics of N passing through x are also geodesics of M. Submanifolds geodesic at x preserve distances to x. This is a vital property of PGA, since variance is defined as the average squared distance to the mean, μ. The goal of PGA is to find a sequence of (nested) geodesic submanifolds centered at the mean. These are the images $S = \text{Exp}_\mu V$ of linear subspaces V of $T_\mu M$. Thus, submanifolds geodesic at the mean can be viewed as the

Intrinsic Mean for a set of points in S²
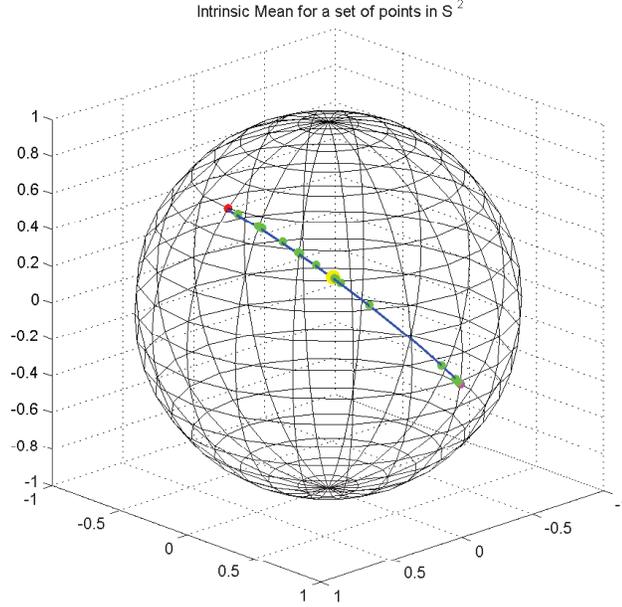
Figure 4.3: Intrinsic mean (yellow) of a set of random points generated on a geodesic. IM is calculated by solving the minimization problem of equation (4.1Principal Geodesic Analysisequation.4.1.1).

generalization of linear subspaces in PCA.

PGA requires us to project the data onto a lower dimensional geodesic submanifold N of M. The point on N nearest to x in Riemannian distance is defined by the projection map $\pi_N : M \to N$.

$$\pi_N(x) = \operatorname*{argmin}_{y \in N} d_R(x, y)^2, \tag{4.4}$$

where $d_R(x, y)^2$ denotes the usual Riemannian squared distance. Depending on the manifold, $\pi_N(x)$ might be difficult to compute. Therefore, we can approximate it by linearizing the manifold, i.e., the data is projected onto the tangent space $T_\mu M$ using the *log* map and regular PCA is performed. In an equivalent way, PGA extends PCA by finding geodesic subspaces in which the variance is maximized. More on computing the projection operator can be found in [19], and [45] where PGA is applied to

medical imaging data.

We give an example by applying PGA on a set of points on the 2-sphere. We uniformly generate some points on a geodesic and then add some noise to perturb them. The intuition is as follows. Since we know the original points were generated on a geodesic, we would expect that the principal geodesic with the heaviest weight in variance to follow the same path as our original geodesic. Figure 4.4 (a) displays two sets of data sampled on the original geodesic. Green points are generated uniformly on the geodesic and black points have noise added to them. Figure 4.4 (b) shows the eigenvectors projected onto the tangent plane about the IM and (c) using the exponential map to project them back onto the manifold and solve for the new geodesics. As suggested earlier, our main principal geodesic is in line with the original one thus, accounting for the majority of the variation in the data. The results are recorded in Table 4.1

Table 4.1: Summary of PGA.

PC	Var. Expl(%)
1	78.34
2	21.66

PGA gives us the ability to perform dimension reduction analysis on manifold data. Analogous to PCA, we find a sequence of nested geodesic submanifolds which best represents the data. The IVS is a 2-dimensional surface in $\mathbb{R}^3$. It is a differentiable manifold evolving continuously in time. Points on the IVS can then be classified as manifold data. The natural response that arises is what type of manifold can the IVS be classified with, and even more important, how can we apply PGA. In the next section we give the steps required to build a geometric framework of the LVS.
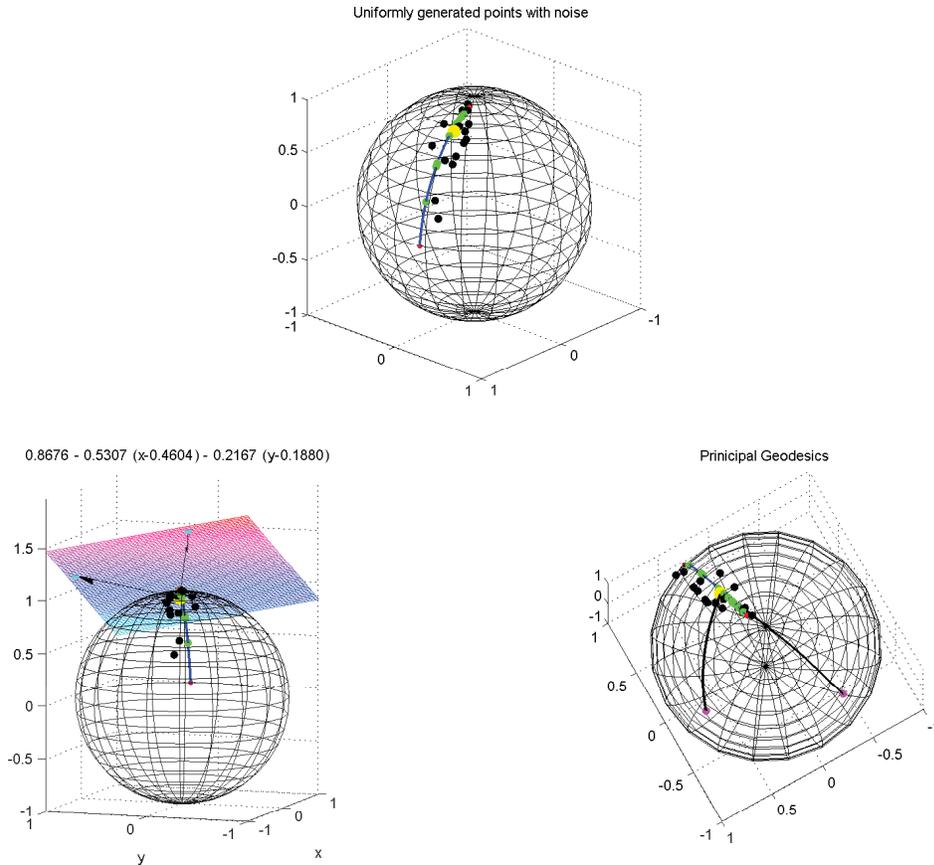
Figure 4.4: From the top going counterclockwise, Original geodesic with uniformly generated points (green), point with added noise (black), and IM (yellow) of the noisy points. Projection of eigenvectors onto the tangent plane about the IM. Using the exponential map to project back onto the sphere and calculate the principal geodesics (black lines).

4.2 Manifold representation of the volatility surface

Quantifying the smile or skew in a coherent model is of great importance, especially for hedging purposes. The development of local volatility models (LVM) has given us the ability to perform such a task. LVMs are self-consistent, arbitrage-free, and can be calibrated to match observed market prices (see Chapter 1). The LVS can be seen as a random surface evolving in low-dimensional manifold of surfaces. This manifold has a given topology from which we can extract various properties. What makes this

a unique type of manifold can be seen by the following example. Take a point x on the sphere $S^2(r) \in \mathbb{R}^3$. Now everyday produce the same sphere with the same point but with a different radius each time. The only dependance on the point x is due to the change in radius. Essentially, a sphere is a sphere. As long as we know the radius then there is no problem relocating the point x day after day. Unlike the LVS or even the IVS, for every new daily set of option prices, we get a new surface. The dynamics of this surface depend on a variety of factors such as; no-arbitrage conditions, the regression coefficients just to name a few. Thus, a point z on the LVS today, may not so easily be located on the surface generated given tomorrow's data.

One way around this problem is to try and identify the space in which the LVS lives in. This 2-dimensional manifold lives in the space of all possible 2-dimensional manifolds with some "special" characteristics. Recall when we originally smoothed the surface we were restricting ourselves to a finite grid both in the moneyness and time to maturity direction. Intuitively we can think of the LVS as a manifold evolving infinitely through moneyness and maturity. Using a parametric representation of the IVS we can rewrite $\hat{\sigma}$ as

$$\hat{\sigma}(\kappa, \tau) = \alpha_0 + \alpha_1 \kappa + \alpha_2 \kappa^2 + \alpha_3 \tau + \alpha_4 \tau^2 + \alpha_5 \kappa \tau, \qquad (4.5)$$

where the coefficients α_i are easily solved by two-dimensional least-squares.

It is worth noting that the LVS generated by this interpolation is not arbitrage free and may not even respect the slope bounds. We are forcing this polynomial approach which guarantees us that $\hat{\sigma}$ is $\mathcal{C}^{1,2}$ in both K and τ. More complex optimization criteria is required to produce a surface free of the types of arbitrage mentioned earlier in Chapter 2. Our choice of order 2 polynomial is motivated by similar reasoning as

in Chapter 2. Not much is gained by expressing it in higher order although, the full representation would be an infinite dimensional polynomial.

We hope to apply PGA on the on the IVS or better yet the LVS because of its PDE description and no arbitrage conditions. For a given range of moneyness and time to maturity, define $\mathbf{\Sigma}$ as our data set

$$\mathbf{\Sigma} = \{\sigma_1(\kappa, \tau), \sigma_2(\kappa, \tau), \ldots, \sigma_n(\kappa, \tau)\}, \tag{4.6}$$

where the subscript denotes the local volatility surface generated on day i. Following the works of Fletcher et al. [19], we then would need to define the space in which $\mathbf{\Sigma}$ lives in. If we parameterize the LVS using equation (4.5Manifold representation of the volatility surfaceequation.4.2.5), then rather naively we can define $\mathcal{P}_n$, as the space of all second order two-dimensional polynomials with rational coefficients with dynamics $\sigma(\kappa, \tau)$

$$\mathcal{P}_n = \{\sigma(\kappa, \tau) : \kappa > 0, \tau > 0, \alpha_i > 0, \alpha \in \mathbb{Q}\}. \tag{4.7}$$

Once we have found the space that best represents the geometry of the LVS then we now have a foundation on which PGA can be implemented. This manifold induces a metric g^* which would enable us to calculate distances between points on $\sigma(\kappa, \tau)$. Then by definition (A.6defis.A.6), this surface constitutes a Riemannian manifold defined by $\mathbf{\Sigma}^*$. Using a parametric representation of the surface like the one given by (4.5Manifold representation of the volatility surfaceequation.4.2.5) the metric g^* can be calculated by

$$g^* = \begin{pmatrix} 1 + \hat{\sigma}_\kappa(\kappa, \tau)^2 & \hat{\sigma}_\kappa(\kappa, \tau)\hat{\sigma}_\tau(\kappa, \tau) \\ \hat{\sigma}_\tau(\kappa, \tau)\hat{\sigma}_\kappa(\kappa, \tau) & 1 + \hat{\sigma}_\tau(\kappa, \tau)^2 \end{pmatrix}$$

Once we have the metric, we can then define a distance and find the geodesics by solving the Euler-Lagrange equations. The geodesics will enable us to find the *Exp* map which brings us from the tangent plane to the surface and back via the *Log* map. We are then able to define the intrinsic mean and apply PGA to data taking values on the LVS.

Unlike PCA and CPC, we do not want to restrict our analysis to slices alone. Ultimately we would like to work with a time series of surfaces as the data as opposed to slices or surface points. Algorithm 1 would have to be modified in such a way where each observation x_i would correspond to a surface $\sigma_i(\kappa, \tau)$. Hence, the IM would give us the average volatility surface, $\overline{\sigma(\kappa, \tau)}$. Applying PGA to this data would gives the principal geodesic submanifolds. These would be the k most prominent submanifolds that best describe the variance of the LVS over time. However, this program is beyond the scope of this thesis.

Chapter 5

Conclusion and Future Research

The implied volatility surface is a very strike and time to maturity dependant structure. Quoting an IVS on a particular day is synonymous with specifying prices for all call and put options on that date. We have provided an in-depth statistical analysis of the various properties defining the surface. We find that the IVS attains its minimum near the ATM level and its term structure decreases as time to maturity increases. We also introduced the theory of local volatility and showed how the LVS can be expressed in terms of the IVS. We described the various types of arbitrage violations encountered when manipulating the IV data. Producing an IVS free of arbitrage is in general not possible as suggested by the literature. We presented various interpolation methods for smoothing the surface, and ultimately chose to use a local polynomial smoother for its effectiveness in capturing the overall dynamics of the surface. Using end-of-day European option prices obtained from the S&P500 Index we reverse engineered the BSM formula to solve for the IVs. Once a surface was fitted to the raw IVs we moved onto describing the components which best describe the variance of the IVS through the use of dimension reduction techniques. We applied PCA and CPC to slices of the IVS for a fixed time to maturity. As is observed in the literature, we found that the IVS can best be described using 3 principal components. The level,

slope and twist components. With these 3 PCs we can build a lower dimensional factor model of the surface.

Dimension reduction techniques such as PCA and CPC rely on the assumption that the data is linear. That is, it assumes that the data is sampled from a vector space. The IVS is a 2-dimensional surface in $\mathbb{R}^3$. It is a continuously differentiable manifold evolving continuously in time. This motivates us to use PGA because of it's non linear dimension reduction technique. PGA uses the inherent properties of the manifold such as the mean and variance, to produce a lower dimensional factor model. What distinguishes PGA from other tractional reduction methods is its ability to manipulate manifold data. We gave an example by applying PGA to a set of points on the 2-sphere.

Although we couldn't apply PGA to the volatility surface since we could not attribute a geometrical representation, we never the less outline a procedure for future research. Following in the steps of Fletcher et al. [19] we strive to study the geometric properties of the LVS as outlined in the end of Chapter 4. We aim at applying PGA on actual surface data as opposed to slices as was done in Chapter 3. This will enable us to capture the entire dynamics of the surface simultaneously across all moneyness and time to maturity. Therefore, producing a geodesic submanifold factor model which best describes the surface.

Appendix A

Differential Geometry

We briefly introduce some notions of differential geometry mostly present in Chapter 4. Our attention will mainly be focused on Riemannian manifolds and the statistics on such surfaces. For a good introductory read on differential geometry we recommend the book by O'Neil [36]. For a more sophisticated read of Riemannian Geometry and its applications we defer the reader to [33], [30], and [37]. In the simplest terms, we begin our adventure with the notion of a surface. Always keeping in the back of our minds the link between the IVS. A surface defined in its simplest terms is a set M, a collection of any objects. An abstract patch in M is a 1-1 function $x : D \rightarrow M$ from an open set D of $\mathbb{R}^2$ into the set M.

Definition A.1. A surface can be described as a set M equipped with a collection P of abstract patches in M.

The following definition is taken from [36].

Definition A.2. An n-dimensional differentiable manifold M is a set furnished with a collection P of abstract patches (one-to-one functions $x : D \rightarrow M, D$ an open set in $\mathbb{R}^n$) satisfying

1. The covering property: The images of the patches cover M.

2. For any patches x, y in P, $y^{-1}x$ and $x^{-1}y$ are Euclidean differentiable.

3. For any two points $p \neq q$ in M there are disjoint patches x and y with p in $x(D)$ and q in $y(E)$.

Lemma A.1. A surface M is said to be compact if and only if it can be covered by the images of a finite number of 2-segments (differentiable maps) in M.

Definition A.3. Let p and q be points of $M \subset \mathbb{R}^3$ and consider all curves α in M connecting points p to q. The intrinsic distance $d_I(p, q)$ in M is the greatest lower bound of the lengths $L(\alpha)$ of these curve segments.

$$L(\alpha) = \inf_{\alpha \in M} \{d_{I1}(p, q), d_{I2}(p, q), \ldots\}.$$

Definition A.4. An isometry $F : M \to \overline{M}$ of surfaces in $\mathbb{R}^3$ is a 1-1 mapping of M onto $\overline{M}$ that preserves dot products of tangent vectors.

An important consequence of the above definition is that isometries preserve distances. If $F : M \to \overline{M}$ is an isometry of surfaces in $\mathbb{R}^3$ then

$$d_I(p, q) = \overline{d_I}(F(p), F(q)) \quad \forall p, q \in M.$$

Let us now quickly pass over these definition to Riemannian geometry.

Definition A.5. A Geometric Surface is an abstract surface M furnished with an inner product $<,>$ on each of it's tangent planes.

This definition resembles that of a regular Euclidean surface given by definition (A.1defis.A.1). The addition here is the geometric structure provided by the collection of all these inner products can be described as a metric tensor, g on M.

We can think of this as a smooth grouping of inner products on the manifold M.

Definition A.6. A Riemannian Metric on a smooth manifold is a 2-tensor field g that is symmetric $g(X, Y) = g(Y, X)$ and positive definite, $g(X, Y) > 0$ for $X \neq 0$. A Metric Tensor is a function on all ordered pairs of tangent vectors v, w at the points p of M on each tangent space $T_p M$ which changes "differentially" with the point p . In short it is a method to define distance on a manifold. It is usually written as

$$g_p < v, w >=< v, w >_p$$

The above definitions can be combined in the following way

$$A \ Surface \ + Metric \ Tensor = Geometric \ Surface$$

$$A \ Manifold \ + Metric \ Tensor = Riemannian \ Manifold$$

Definition A.7. A curve γ in $M^n \subset \mathbb{R}^{n+1}$ is a called a geodesic of M^n if its acceleration γ'' is always orthogonal to M^n. Furthermore γ has constant speed,

$$(\|\gamma'\|^2)' = 2\gamma' \cdot \gamma'' = 0.$$

Geodesics can be thought of as the manifold generalization of a straight line. Geodesics play a fundamental role in the study of Riemannian geometry much in the same way straight lines are so crucial in Euclidean geometry. The smallest geodesic connecting two points in M is called a minimizing geodesic as displayed in Figure 4.1 of Chapter 4. Hence, the notion of distance can now be formulated. For points $p, q \in M$, we define the space

$$\Gamma_{p,q} = \{\gamma : [0, 1] \to M : \gamma \text{ is piecewise } \mathcal{C}^\infty \text{ and } \gamma(0) = p, \gamma(1) = q\}.$$

Now the Riemannian distance $d_R(p, q)$ between points p and q in M is defined as:

$$d_R(p, q) = \inf_{\gamma \in \Gamma_{p,q}} \{\mathcal{L}(\gamma)\}$$

where $\mathcal{L}(\gamma)$ is the usual formula for arclength. Two properties need to be presented here. One being the notion of completeness, and the other defining what is meant by a minimizing geodesic.

Definition A.8. A Riemannian manifold where all geodesics γ exist for all time is called *geodesically complete*.

Another way of rephrasing Definition A.8 is by saying that a manifold M is complete if all geodesics extend indefinitely. The next theorem links these two properties together. for all compact and non-compact manifolds.

Theorem A.1. (Hopf-Rinow) If M is a complete manifold, then any two points p and q in M can be joined by a geodesic of length $d_R(p, q)$. Defining $d_R(p, q)$ as in definition (A.7defis.A.7), p and q can be joined by a minimizing geodesic.

For $\mathbb{R}^n$, a straight line can extend to eternity and for any two points, the Hopf-Rinow Theorem guarantees the existence of a minimizing geodesic, namely the unique straight line segment $\overline{PQ}$. For $M = S^2$, the minimizing geodesic is the shortest arc of the great circle connecting the two points. Uniqueness for the sphere however, does not hold if p and q are located at opposite poles.

We also have another tool which helps us to define minimizing geodesics on M. Given points p and q in M, let v belong to the tangent plane $T_p M$, the Exponential Map denoted $Exp_p(v)$ is a map which constructs geodesics.

Theorem A.2. Let M be the usual Riemannian manifold, with $p \in M$, and $v \in T_pM$. Then for some small $\epsilon > 0$, there is only one geodesic, $\overline{\gamma}$ such that:

$$\overline{\gamma} : [0, \epsilon] \to M$$

with initial conditions $\overline{\gamma}(0) = p$, $\overline{\gamma'}0) = v$.

This theorem implies the uniqueness and existence of geodesics. We can now state a proper definition for the exponential map.

Definition A.9. For $p \in M$, define the space V_p as

$$V_p := \{v \in T_pM : \gamma_v \text{ is defined on } [0,1]\}$$

The exponential map is defined by

$$Exp_p : V_p \to M$$

$$v \mapsto \gamma_v(1).$$

The exponential map gives us a link between the tangent map and the manifold.

Theorem A.3. The Exponential map Exp_p maps a neighborhood of $0 \in T_pM$ diffeomorphically onto a neighborhood of $p \in M$.

There exists some diffeomorphism say Φ, which brings us from the tangent plane T_pM to the manifold M and **vice versa**. The inverse exponential map, is given by the Log map. By the above theorem there exists a neighborhood U of p which is mapped by Log_p diffeomorphically onto a neighborhood $0 \in T_pM$. Hence we have $Log_p : Exp_p(U) \to T_pM$. We now have a way of going back and forth from one space to the other. This property will be the key factor in finding the IM on a manifold.

Next, we introduce some notions on Lie groups and the role they play in this thesis.

Definition A.10. A *Lie group* G is a differentiable manifold endowed with a group structure such that,

$$(x, y) \mapsto xy \quad : G \times G \to G$$
$$x \mapsto x^{-1} : G \to G,$$

are group operations compatible with the differentiable structure.

Lie groups can be used to describe transformations of smooth manifolds. Lie algebras are constructed by linearizing Lie groups.

A Riemannian symmetric space is a connected manifold M such that at each point $x \in M$ the mapping that reverses geodesics at that point is an isometry (distance preserving). The Euclidean space $\mathbb{R}^n$ is a symmetric space, so are spheres, $\mathbb{S}^n$, and hyperbolic spaces, $\mathbb{H}^n$. Ultimately symmetric spaces offer us a way for computing geodesics while using the Lie group actions inherent on those manifolds.

Definition A.11. A Riemannian manifold M is called symmetric if for every $x \in M$ there exists an isometry $\varphi : M \to M$ such that

$$\varphi_x(x) = x$$
$$D\varphi_x(x) = -Id.$$

A subgroup $H \subset G$ which permutes with all operations in G is called an **invariant subgroup** of G. Isomorphic groups are almost identical at the algebraic level. Thus, when an isomorphism ϕ exists, between a group and a matrix group, it is often more convenient to study the matrix representation of the group since matrix properties are so familiar. For $H \subset G$, we can write every group element in G a a product of an element h in its subgroup H with a group element in a "quotient" or *coset* denoted G/H.

The space of all such cosets is a smooth manifold. Furthermore, let M be a symmetric space and define an arbitrary base point $p \in M$. M can be written as a homogeneous space $M = G/H$, where G is a connected group of isometries, and the isotropy subgroup H is compact. It is interesting to note that M need not be compact yet can still be represented by a compact group. Geodesics on M can be computed through the group action. They are the image of the action of a one-parameter subgroup of G acting on the base point p.

Bibliography

[1] C. Alexander. Principal Comoponent Analysis of Volatility Smiles and Skews. *The University of Reading, Working Paper*, 2001.

[2] J. Anderson and B. Huge. Volatility Interpolation. *RISK*, 2011.

[3] T. Ané and C. Labidi. Implied Volatility Surfaces and Market Activity over Time. *Journal of Economics and Finance*, 25, 2001.

[4] M. Benko. *Functional Data Analysis with Applications in Finance*. PhD thesis, Humboldt-Universität zu Berlin, 2006.

[5] F. Black and M. Scholes. The Pricing of Options and Coroporate Liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[6] A.W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach With S-Plus Illustrations*. Oxford University Press, 1997.

[7] D.T. Breeden and R.H. Litzenberger. Prices of state-contigent claims implicit in option prices. *The Journal of Business*, 51:621–651, 1978.

[8] W.S Cleveland and C. Loader. *Smoothing by Local Regression: Principles and Methods*. Springer-Verlag, 1996.

[9] R. Cont and José. da Foncesca. The Dynamics of Implied Volatility Surfaces. *Quantitative Finance*, 2:45–60, 2002.

[10] S. Crepey. Calibration of the local volatility in a generalized Black-Scholes model using Tikhonov regularization. *Journal on Mathematical Analysis, volume =*.

[11] G. Debreu. Theory of Value: An axiomatic analysis of economic equilibrium. *Yale University Press*, 1959.

[12] E. Derman and I. Kani. The Volatility Smile and Its Implied Tree. *Goldman Sachs Quantitative Strategies Research Notes*, 2:45–60, January 1994.

[13] E. Derman, L. Kani, and N. Chriss. Implied Trinomial Trees of the Volatility Smile. *Goldman Sachs Quantitative Strategies Research Notes*, 1996.

[14] B. Dupire. Pricing and Hedging with Smiles. *Paribas Capital Markets Swaps and Options Research Team*, 1993.

[15] M.R. Fengler. Arbitrage-free Smoothing of the Implied Volatility Surface. *SFB 649 Discussion Paper, Humbolt-Universität zu Berlin*, 2005.

[16] M.R. Fengler. A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics. *SFB 649 Discussion Paper, Humbolt-Universität zu Berlin*, 2005.

[17] M.R. Fengler. *Semiparametric Modeling of Implied Volatility*. Springer Finance, 2005.

[18] M.R. Fengler and W.K. Hardle. The Dynamics of Implied Volatilities: A Common Principal Components Approach. *Review of Derivatives Research.*, 6:179–202, 2003.

[19] P.T Fletcher, C. Lu, S.M Pizer, and S. Joshi. Principal Geodesic Analysis for the Study of Nonlinear Statistics of Shape. *IEEE Trans Med Imaging*, 23, 2004.

[20] B. Flury. *Common Prinicipal Components and Related Multivariate Models*. John Wiley and Sons, 1988.

[21] B. K. Flury and J.P. Airoldi. An application of common prinicipal component analysis to cranial morphometry of Microtus californicus and M.ochrogaster(Mammalia,Rodentia). *Journal of Zoology*, 216, 1988.

[22] K.R. French and R. Roll. Stock Return Variances: The Arrival of Information and the Reaction of Traders. *Journal on Financial Economics, volume =.*

[23] J. Gatheral. *The Volatility Surface A Paractitioner's Guide.* John Wiley and Sons, 1988.

[24] G. Grimmett and D. Stirzaker. *Probability and Random Processes.* Oxford University Press, 2001.

[25] G. Grith, W.K Härdle, and M. Schienle. Nonparametric Estimation of Risk-Neutral Densities. *SFB 649 Discussion Paper, Humbolt-Universität zu Berlin*, 2010.

[26] C. Harvey and R. Whaley. S&P 100 Index Option Volatility. *Journal of Finance*, 46, 1991.

[27] D. Higham. *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation.* Cambridge University Press, 2004.

[28] J. Hull, T/ Daglish, and W. Suo. Volatility Surfaces: Theory, Rules of Thumb, and Empirical Evidence. *Quantitative Finance*, 7:507–524, May 2007.

[29] E. Jondeau, S-H. Poon, and M. Rockinger. *Financial Modeling Under Non-Gaussian Distributions.* Springer Finance, 2007.

[30] J. Jost. *Riemannian Geometry and Geometric Analysis.* Springer, Universitext, 2008.

[31] N. Kahalé. An Arbitrage-free Interpolation of Volatilities. *Risk Magazine*, 17: 102–106, May 2004.

[32] N. Kahalé. Smile interpolation and calibration of the local volatility model. *ESCP-EAP*, 2005.

[33] J.M. Lee. *Riemannian Manifolds: An Introduction to Curvature.* Springer, Graduate Texts in Mathematics, 1997.

[34] R.W. Lee. *Implied Volatility: Statics, Dynamics, and Probabilistic Interpretation.* Recent Advances in Applied Probability, Springer, 2002.

[35] R.C. Merton. Theory of Rational Option Pricing. *The Bell Journal of Economics and Management Science*, 4, 1973.

[36] B. O'Neil. *Elementary Differential Geometry.* Academic Press, 2006.

[37] P. Peterson. *Riemannian Geometry.* Springer, Graduate Texts in Mathematics, 2006.

[38] R. Rebonato. *Volatility and Correlation.* Wiley Series in Financial Engineering, John Wiley Sons Ltd., 1999.

[39] F. Rouah. Uisng the Risk Neutral Density to Verify No Arbitrage in Implied Volatility. URL `http://www.frouah.com/pages/finmath.html`.

[40] D. Rupert and M.P. Wand. Multivariate Locally Weigthed Least Squares Regression. *The Annals of Statistics*, 22, 1994.

[41] D. Rupert, S.J. Sheather, and M.P. Wand. An Effective Bandwidth Selector for Local Least Squares Regresson. *Journal of the American Statistical Association*, 90:1257–1270, 1995.

[42] D. Shimko. Bounds on Probability. *RISK*, 6:33–37, 1993.

[43] S.E. Shreve. *Stochastic Calculus for Finance II Continuous-Time Models.* Springer Finance, 2000.

[44] G. Skiadopoulos, S. Hodges, and L. Clewlow. The Dynamics of the S&P500 Implied Volatility Surface. *Review of Derivatives Research*, 3:263–282, 1999.

[45] S. Sommer, F. Lauze, S. Hauberg, and M. Niehen. Manifold Valued Statistics, Exact Prinicipal Geodesic Analysis and the Effect of Linear Approximation. *Computer Vision- ECCV 2010*, 6, 2010.

[46] Y. Wang, H. Yin, and L. Qi. No-Arbitrage Interpolation of the Option Price Function and its Reformulation. *Journal of Optimization Theory and Applications*, 120:627–649, 2004.

[47] P. Wilmott. *Paul Wilmott on Quantitative Finance 3 Volume Set*. Wiley, 2006.

[48] G. Woldberg and I. Alfy. An Energy-Minization framework for monotonic cubic spline interpolation. *Journal of Computational and Applied Mathematics*, 143: 145–188, 2002.