SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Volatility Smirk
Reconstruction of the Unknown Volatility Curve

av

Johan Cassel

2012 - No 3

Volatility Smirk
Reconstruction of the Unknown Volatility Curve

Johan Cassel

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Sara Maad Sasane

2012

Acknowledgements

I would like to thank Sara Maad Sasane for allowing me to do this work despite my lack of prior knowledge in the field, as well as for the help during the writing process. I also want to thank Annemarie Luger, my examiner, for valuable comments.

A special thanks to my friends Benjamin, Cecilia and Björn for teaching me how Matlab works and to Josefin, Disa and Fia for helping me correct mistakes. Finally, thanks to Johan and Jan for making long days of writing more enjoyable and to Bea for putting up with me.

Contents

1 Introduction

In 1973, Fischer Black and Myron Scholes published a paper on how to price an option. They did this by replicating an option via a portfolio consisting of just the right amount of bonds and stocks.

From this portfolio they derived a partial differential equation that determines the option price. That is, the only price which, given their model of stock price movement, ensures that one cannot make a sure profit by constructing an appropriate portfolio. More than that, they derived a partial differential equation that lead to an explicit solution with only a single variable not directly observable from the market. This lead to tremendous effects in the trading of options and a better understanding on how to price financial derivatives.

Robert C. Merton published a paper that increased the mathematical understanding of the model, as well as proving that some of the assumptions made in the paper published 1973 were redundant. For their contribution, Merton and Scholes were awarded the 1997 Nobel Price in Economics (The Sveriges Riksbank Price in Economic Sciences in Memory of Alfred Nobel). Fischer Black died in 1995, and was thus ineligible for the price. He was however mentioned as a contributor by the Royal Swedish Academy of Science.

Although a breakthrough, the model has turned out to not describe the reality perfectly and several suggestions on how to improve their model have been suggested. Many of these alterations lead to partial differential equations which cannot be solved explicitly.

The most well known discrepancy in the model from the reality is called the *volatility smile*. It refers to their assumption that the volatility of a stock is constant. If one reverts their formula with observed market prices and solves it for the volatility then one will see that the volatility is not constant, and in some cases it appears in the shape of a smile, hence the name. For stocks, the volatility smile often appear as a smirk.

The aim of this thesis is to review one of the suggested models that allows for a volatility function. Given this model, we try to reconstruct a volatility function from observable market prices. From this model we derive the partial differential equation describing an option price and a numerical method for solving the differential equation is discussed. More than that, it is shown how one can use splines in approximating the volatility function and a scheme to reconstruct the volatility function, so that the model matches observed market prices, is given.

2 Stochastic definitions and Itô's formula

We will work with a model of stock price behaviour. Normally stock prices act in a seemingly random way, with many small steps away from a general trajectory, if such a trajectory exists at all. Therefore, the model that is usually used is using a special kind of stochastic process called a "*Wiener process*" or "*Browian motion*", the model is a stochastic differential.

We here state a few definitions and the stochastic version of the *chain rule* in order to gain some understanding about the model we work with. This part is very brief, and to get a better understanding one can read for example [5] or [13]. All definitions originate from these two works.

2.1 Basic definitions in probability theory

Definition. Let Ω be a set. $\mathcal{F}$ is a σ-algebra of subsets of Ω if

(a) $\emptyset, \Omega \in \mathcal{F}$.

(b) If $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$ where $A^c := \Omega \setminus A$.

(c) If $A_1, A_2, \cdots \in \mathcal{F}$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$.

Definition. Let Ω be a non-empty set. We say that a function $P : \Omega \to [0, 1]$ is a *probability measure* if it has the following properties:

(a) $P(\emptyset) = 0, P(\Omega) = 1$.

(b) If $A_1, A_2, \cdots \in \mathcal{F}$, then $P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k)$.

(c) If $A_1, A_2, \cdots \in \mathcal{F}$ are disjoint, then $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

We refer to a set $A \subseteq \mathcal{F}$ as an *event*, and a point $\omega \in \Omega$ as a *sample point*. We say that $P(A)$ is the *probability* of the event A. If we have two events, A,B $\in \mathcal{F}$ then they are said to be independent if $P(A \cap B) = P(A)P(B)$.

Definition. If we take a set Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω and a probability measure P defined on Ω, we say that the triple $(\Omega, \mathcal{F}, P)$ is a *probability space*.

Now, the set Ω could for example consist of only *heads* or *tails*, in case of a coin toss. We could of course observe the outcome of a coin toss, and

write it down as either 'heads' or 'tails'. However we are generally more interested in some measurable quantity. This quote from [13] (p. 106) is a motivation to the introduction of random variables,

"*We can think of the probability space as being an essential mathematical construct, which is nevertheless not "directly observable". We are therefore interested in introducing mappings X from Ω to $\mathbb{R}$, the values of which we can observe. These mappings are called random variables.*"

Definition. We say that a mapping $X : \Omega \to \mathbb{R}$ is a *random variable* if, $\forall t \in \mathbb{R}$,

$$\{\omega | X(\omega) \leq t\} \in \mathcal{F}.$$

In the example of the coin toss we could for example have a random variable that gives us the value 2 in case of heads, and 0 in case of tails. This random variable then represents the outcome of betting on the result of a coin toss where you get twice the money you bet if heads comes up. Normally one does not display the random variables dependence on the sample point $\omega \in \Omega$.

Definition. If we have two random variables X, Y, then they are *independent* if, for all $t, s \in \mathbb{R}$,

$$P(X \leq t \text{ and } Y \leq s) = P(X \leq t)P(Y \leq s).$$

We interpret this as that the two random variables convey no information about the other one. That is, if two random variables are independent and we receive information of one of the two, the probability distribution of the other one is still the same.

Definition. If X is a random variable, defined on some probability space $(\Omega, \mathbb{F}, P)$, the *expected value* of X is

$$E[X] := \int_\Omega X dP.$$

We define the *variance* of X as

$$\text{Var}(X) := E[(X - E(X))^2] = E[X^2] - (E[X])^2.$$

Definition. A random variable X is called *normal* with mean μ and variance σ^2 if, for all $-\infty \leq a < b \leq \infty$,

$$P(a \leq X \leq b) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_\Omega e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

We write this short as "X is $N(\mu, \sigma^2)$".

Definition. (i) If we have a collection of random variables $X(t),\ 0 \leq t < \infty$, all defined on the same probability space $(\Omega, \mathcal{F}, P)$, we say that this collection is a *stochastic process*.

(ii) For each point $\omega \in \Omega$ we have a corresponding *sample path* that is the mapping $t \mapsto X(t, \omega)$.

If we run an experiment and observe the random values of X over time, we are looking at a sample path $\{X(t, w)|\ t \geq 0\}$ for some fixed $\omega \in \Omega$. Thus a stochastic process can either be viewed as a collection of random variables or as a random function (path) over time. This path will generally be different every time we run the experiment.

We are now ready to define the special kind of stochastic process we are mainly interested in.

Definition. A stochastic process $W(t)$ is called a *Wiener process* or *Brownian motion* if

1. $W(0) = 0$.
2. Each sample path is continuous.
3. $W(t) - W(s)$ is $N(0, t - s)$ for all $t > s$, that is, normal with mean μ and variance $t - s$.
4. For all distinct choices of times $0 < t_1 < t_2 < \cdots < t_n$ the increments $W(t_1), W(t_2) - W(t_1), ..., W(t_k) - W(t_{k-1})$ are independent random variables.

Property 4 tells us that the increments between two time steps does not depend on previous increments.

2.2 Stochastic integrals, stochastic differentials and Itô's formula

We are soon going to introduce a stochastic differential, but in order to do so, we need the Itô stochastic integral. However, constructing it goes beyond the scope of the paper. One can read more about it in chapter 4 of [5]. I will however state a few properties of it that will play a part in this paper.

The aim is to have a integral for a certain class of stochastic processes. This class should be *nonanticipating*, which means that it depends only on

information available up until time t. Given a Brownian motion W one can define the Itô stochastic integral

$$\int_0^t Y dW$$

for stochastic processes Y with the property that at each time $0 \leq s \leq t$, $Y(s)$ *depends only on* $W(\tau)$ *for times* $0 \leq \tau \leq s$ *but not on times* $s \leq \tau$. This means that Y is a nonanticipating stochastic process; at every time it only depends on the Brownian motion up until that time, but not the future movement. As already noted, we will not look further into the Itô integral here.

Definition. Suppose that X is a stochastic process satisfying

$$X(t_2) - X(t_1) = \int_{t_1}^{t_2} F dt + \int_{t_1}^{t_2} G dW \tag{1}$$

for some $F, G, \in \mathbb{L}^2(0,T)$ and all times $0 \leq t_1 \leq t_2 \leq T$. Then we say that X has the stochastic differential

$$dX = F dt + G dW. \tag{2}$$

One needs to note that this is just a shorter way of typing the difference between $X(t_2)$ and $X(t_1)$, strictly speaking, "dX", "dt" and "dW" have no meaning alone. Of the two integrals in (1), the left one is an ordinary integral whereas the right one is a stochastic integral.

The model we are going to work with later will be on the form of a stochastic differential. In order to manipulate it we are going to use the following formula, which we can think of as the counterpart to the ordinary *chain rule* for stochastic differentiation.

Theorem 1 (Itô's formula)

Suppose that X *has a stochastic differential*

$$dX = F dt + G dW$$

for $F, G \in \mathbb{L}^2(0,T)$. *Let* $u : \mathbb{R} \times [0,T] \to \mathbb{R}$ *be continuous and suppose that* $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$, $\frac{\partial^2 u}{\partial x^2}$ *exist and are continuous. Let*

$$Y(t) := u(X(t), t),$$

then Y has the stochastic differential

$$\begin{aligned} dY &= \frac{\partial u}{\partial t}dt + \frac{\partial u}{\partial x}dX + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2 dt \\ &= \left(\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x}F + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}G^2\right)dt + \frac{\partial u}{\partial x}G dW. \end{aligned} \tag{3}$$

We will not prove this theorem here, a proof can be found in for example [5]. We will however use the theorem when deriving the partial differential equation that gives us a correct option price.

3 How to price an option

In order to work with options we have to describe what an option is and how we should price them. This section is devoted to explaining what options are and what we wish to achieve when pricing them.

3.1 What is an option?

An option is a derivative financial instrument that specifies a contract between two parties on some underlying asset, for example a stock. The buyer of the option pays for the right to engage in a transaction according to the contract but does not have the obligation to do so. The seller on the other hand is obliged to fulfil the contract if the buyer wishes to do so.

In other words one could say that an option gives you the right to buy (or sell) an underlying asset at a fixed price if that gains you. This fixed price is called the *exercise price* or the *strike price* of the option.

There are different kinds of options. We have *call* options which gives the buyer the right to buy the specified underlying asset at the strike price. We also have *put* options where you instead gain the right to sell the underlying security at the strike price.

An option always has an expiration date after which it is useless and we call this date the *maturity date*. The time remaining until the contract expires, that is, the time left until the maturity date, will be denoted the *maturity time*. An option that can be exercised only on the maturity date is called a *European option*.

In this paper we will be restricting ourselves to European call options. That is an option which gives one the right to buy the underlying security if one

wishes to do so. It can only be exercised on the expiration date.

3.2 The value of an option at the expiration date

Now that we have defined what an option is, we are able to specify the value of an option at the expiration date.

Since you have the right, but not the obligation, to use the option you will never lose any money at the expiration date if you have bought an option. Take for example that you are the owner of a European call option on a stock with a strike price at 100 kr and the underlying stock is worth 120 kr on the expiration date. Exercising the option would result in buying the stock for 100 kr when the market value is 120 kr. In other words you can (instantaneously) sell the stock for 120 kr and gain 20 kr.

If the stock price on the expiration date would instead be 80 kr, then you could buy the stock cheaper from the market than by exercising your option. Due to the fact that you are not obliged to use the option you will not do so. Thus the option is worth 0 kr for you. If we denote the strike price by K and the stock price at the expiration date by S_T then the value of the call option at the expiration date is equal to

$$\max(S_T - K, 0).$$

3.3 Modelling the stock price behaviour

It is clear from the previous section that the value of the option depends on how the stock behaves. A call option becomes more valuable if the price of the stock increases and a put option becomes more valuable if the price of the stock decreases. Thus it is in place to discuss how we anticipate the *movement* of a stock price.

The following paragraph is based on [8], a book written by John C. Hull.

According to Hull (p. 287), the most widely used model of stock price behaviour is the following. One assumes that we know the expected return μ on the stock and that this return is constant in the sense that in a short time interval Δt the expected increase in a stock S is $\mu S \Delta t$. Since we cannot know for certain whether our return will be exactly μ or not, we also take into account some sort of *error*, or variance σ, on the return that we actually get. σ is usually called the *volatility* of the stock. One usually assumes that this error is given as a percentage of the stock price and that it in a short time interval Δt is equal to $\sigma S dW$, where W is a Wiener process. With

these two assumptions, the general model for stock price behaviour is the stochastic differential

$$dS = \mu S dt + \sigma S dW. \tag{4}$$

When pricing an option one has to start with the model of the stock price behaviour since the option price for European options, for a given strike price, is a function solely of the stock price at the expiration date.

From (4), and with the extra assumptions that are stated in appendix A, one can derive the famous Black-Scholes partial differential equation of an European option price $f(s,t)$

$$\frac{\partial f}{\partial t} = rf - r\frac{\partial f}{\partial s}s - \frac{1}{2}\frac{\partial^2 f}{\partial s^2}\sigma^2 s^2, \tag{5}$$

with the boundary condition

$$f(s,T) = \max(s-K, 0),$$

where

s is the current stock price;

K is the strike price;

r is the riskless rate of interest, compounded continuously;

T is the time to maturity from when the option started and t is the time that has passed since then.

The derivation of (5) will be shown in 4.3. When one assumes constant volatility this partial differential equation can be solved explicitly, and that is one of the reasons why it has become so famous. The explicit solution of the option price $f(s,t)$ is

$$f(s,t) = s\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_1} e^{-\frac{1}{2}y^2}dy - ke^{-r(T-t)}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{d_2} e^{-\frac{1}{2}y^2}dy, \tag{6}$$

where

$$d_1 = \frac{\log(\frac{s}{k}) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}},$$

$$d_2 = \frac{\log(\frac{s}{k}) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}.$$

In this explicit solution, the only thing that is not directly observable is the volatility σ, since even the price can be observed on the market. That means that one can, numerically, solve this equation for σ. Determining the

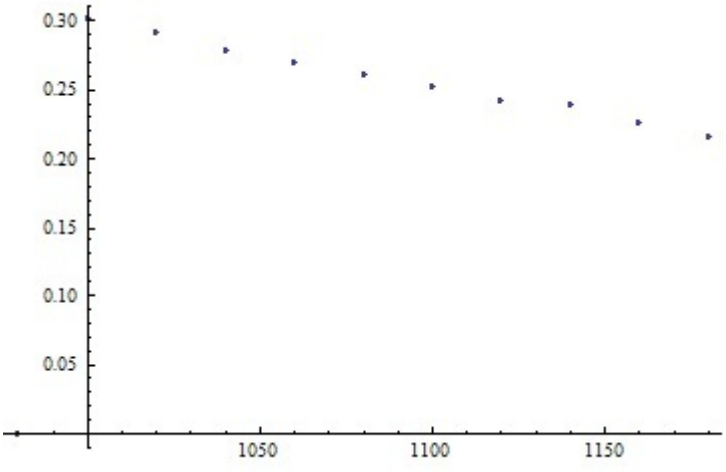
volatility from observed market data is called computing the *implied volatility*.

However, (4) is not widely accepted as a good model for stock price behaviour. Both the assumption that the return on your investment would be constant over time and that the volatility would be constant is far from obvious.

For example, in [4] (p. 107) they say that studies on the implied volatility on underlying securities of options shows that at least the volatility is not constant. If one studies the implied volatility using the BSM formula from actual market prices on options one will notice what is referred to as a *volatility skew*.

3.4 Example of the volatility skew

Here we give an example of the volatility skew. What one means with the volatility skew is related to the assumption made in (4) that volatility is constant. Would that assumption be correct, and we would calculate the implied volatility for different strike prices on the same underlying stock, then the graph of the volatility values plotted against strike prices should be a flat line. According to [4], every volatility yields a unique option price, so we do not need to worry about that. Here we give an example that the implied volatility is not constant over different strike prices. We have taken option prices for *the OMX Stockholm 30 index*. The strike price is on the horizontal axis and the implied volatility on the vertical axis.



It appears as if the volatility would be a decreasing function of the strike price, and this is indeed the general form the volatility skew takes when the

underlying security is a stock [8]. The volatility is supposed to describe how much the stock deviates from the expected path. It does not seem likely that strike prices on options would affect how much the stock deviates from its expected path. It is more likely that the volatility skew represents that the underlying model simply does not describe the reality good enough, rather than saying that the volatility is indeed a function of the strike price.

3.5 The right price of options

What is the goal when pricing an option? What is the *right* price? Since there is a seller and a buyer, we want to price the option such that neither one expects to lose money, that is, if they would calculate the price using the same method. If, say, the one selling the option would expect to lose money *for sure* on the transaction we assume that he or she will not be interested in signing such a contract.

We already know that, at the expiration date, the owner of the option is the only one that can earn money. Thus the buyer must pay the seller a certain price for the option. How much should he pay? Everything the owner of the option gains at the expiration date is a loss for the seller. Therefore it is reasonable to expect that the price of an option would be equivalent with the expected return for the buyer.

However, as they discuss in [1], this is not as intuitive as one would hope. We cannot simply calculate the expected return from a probability distribution and price it that way. In order to do that we would have to have a probability distribution that describes the option return perfect. As mentioned, we want a price that emphasizes that neither the buyer nor the seller expects to lose money *for sure*, and that is the crucial part.

The phrase *to lose money for sure* is related to arbitrage opportunities. An *arbitrage opportunity* is a chance to make money without taking any risk. One example is if you have the chance to sign a contract where you have the right to buy a stock for 100 kr in, say, 10 days. At the same time you sign another contract where your counterpart agrees to buy that same stock from you for 101 kr in 10 days as well. This example guarantees you a net profit of 1 kr, without taking any risk at all.

A reason to expect that opportunities such as this example does not exist is that people (are supposed to) have access to the same information from the market. If you have the chance to sign both contracts in the example at the same time, then so should both of your counterparts and then they will probably engage in a transaction with each other instead of with you. For example, if they agree on the price 100,50 kr, they both gain on that deal

compared to the previous state. We can at least expect such opportunities to be short-lived because if they do exist, many people are going to want to take advantage of them, the demand should make the price rise and thus eliminating this arbitrage opportunity.

Now, how does this make pricing options a bit more complicated? Well, the general idea is that you can construct portfolios, that is gather different securities, financial derivatives and bonds in order to synthesize for example an option. If we are able to construct a portfolio which fully synthesize the option but with a different price than the option, then either the seller or the buyer can gain money without taking a risk, depending of course on whether the portfolio is cheaper or more expensive than the option. The conclusion we make is that if we can construct such portfolio, then there should be an enforcing price of the option that we can derive from this portfolio.

Here we assume that there are no arbitrage opportunities at all. This assumption implies something that is very important. *A risk-free investment cannot give a better effective yield than the risk-free rate of interest*, which is available to everyone.

4 Pricing an option, a derivation of the generalized BSM partial differential equation

Now we know what an option is and what the goal when pricing an option is. In this section we will derive a differential equation describing the price of an option. We will do this with a model that differs slightly from (4).

4.1 Notations and assumptions

Not all assumptions made are mentioned here, but they are listed in appendix A.

As discussed in section 3.3, the assumption that volatility remains constant might not be the best way to model volatility. In this paper we will therefore follow the work of [3] and work with a different model for the underlying security. We assume that the stock price follows a one-factor continuous diffusion model where

$$dS(t) = \mu(S(t), t)S(t) + \sigma(S(t), t)S(t)dW. \tag{7}$$

The previous constant parameters μ and σ are now instead functions of the underlying stock s and the time t. Both μ and σ are assumed to be at least twice continuously differentiable, thus we can apply Itô's formula.

Let $S(t)$ denote the stock price at time t. Let T denote the maturity time (in years) of the option and let K be the strike price. We assume that we know a risk-free interest rate r which is defined such that 1 kr today becomes e^{rt} kr at time t.

We denote the sought value of the option as

$$Y(t) := f(S(t), t).$$

where $f(s, t)$ is the *right*, no-arbitrage, price of the option that we seek given a stock price s and a time t from the starting point.

4.1.1 Interest rates on the Swedish market

We said that we wanted the rate r to be defined such that the growth is e^{rt}, this is the same as saying we have a risk-free rate that gives interest continuously. This is not how interest rates are given by the market. The rates that are assumed to be *risk-free*, for example the Swedish 10-year government bond, are given as effective interest rates where you get paid once every year. We get it on the form we want it by

$$1 + r_{\text{eff}} = e^r \Leftrightarrow$$

$$\ln(1 + r_{\text{eff}}) = r.$$

In this paper we have taken the rate from the Swedish 10-year government bond, and used the formula above to get it expressed on the form we want.

4.2 Setting up the portfolio

The following two sections set up a good portfolio and derives the generalized BSM partial differential equation, that determines the option price. The ideas are based on chapter 8 in [13], written by Amol Sasane. Robert Merton was the one that first derived the equation by creating an appropriate, riskless, portfolio and arguing that the return must be the risk-free return, see [9].

One of the key ideas behind the Black-Scholes-Merton formula is that we can reduce the risk that we take by setting up a portfolio consisting of a position in both the derivative (the option) and the underlying stock. Why is that?

If we own a call option we gain money when the price of the stock goes up but we lose (gain less) money when the price of the stock goes down. If we enter a short position in that stock, i.e. promise that we will sell the stock to a fixed price, we are in the opposite position. We are glad when the

price drops after our sell and we curse ourselves if the price goes up. Thus we may be able to create a portfolio which is irrespective to the movement of the stock price. This is assuming that we can buy and sell parts of stocks which is of course not completely true. When trading in large quantities the error will be small though.

We create a portfolio by buying one call option, we go short in A shares of the underlying stock and invest that money in a bond which has a risk-free interest rate. Let the right, no-arbitrage price of the option be $Y(t)$ and assume that we can express it as $Y(t) = f(X(t), t)$, that is some function of $X(t)$ and t. We say that at time t, we are short $A(t)$ shares of the stock and have $B(t)$ money in bonds. The value $V(t)$ of our portfolio at time t will then be

$$V(t) = Y(t) - A(t)S(t) + B(t)e^{rt}.$$

We want this portfolio to be *self-financing*, which means that we will not have to add or remove money from the portfolio. This can be done in a way that whenever we want to buy or sell stock, we just reallocate money between the stocks and the bond. We thus demand that

$$\forall t, \;\; S(t)dA(t) = e^{rt}dB(t). \tag{8}$$

What does this mean for our portfolio? It means that the value of the portfolio is not changed if we change our allocation of money, the only change must thus come from changes in the underlying asset S(t).

4.3 Derivation of the generalized Black-Scholes-Merton partial differential equation

In 3.3 we showed the partial differential equation describing an option value. Here we give a derivation of it, but with the assumption that a stock follows (7) instead of (4). The change in the assumed model does not affect the derivation of the PDE, however we cannot get an explicit solution of the PDE, and have to resort to numerical solutions.

The value of the portfolio V(t) gives us

$$dV(t) = dY(t) - S(t)dA(t) - A(t)dS(t) + e^{rt}dB(t) + B(t)re^{rt}dt.$$

The self-financing condition (8) cancel out two terms and we get

$$dV(t) = dY(t) - A(t)dS(t) + B(t)re^{rt}dt. \tag{9}$$

Since we assume the option $Y(t) = f(S(t), t)$ we have by *Itô's formula*

$$dY(t) = \frac{\partial f}{\partial t}(S(t),t)dt + \frac{\partial f}{\partial s}(S(t),t)dS(t) + \frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t)dt.$$

Inserting this into (9) yields

$$\begin{aligned} dV(t) = &\frac{\partial f}{\partial t}(S(t),t)dt + \frac{\partial f}{\partial s}(S(t),t)dS(t) + \frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t)dt \\ &- A(t)dS(t) + B(t)re^{rt}dt, \end{aligned}$$

and after rearranging a bit we end up with

$$\begin{aligned} dV(t) = &\left(\frac{\partial f}{\partial t}(S(t),t) + \frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t) + B(t)re^{rt}\right)dt \\ &+ \left(\frac{\partial f}{\partial s}(S(t),t) - A(t)\right)dS(t). \end{aligned} \tag{10}$$

What is interesting here is that we can get rid of the $dS(t)$ part by choosing

$$A(t) = \frac{\partial f}{\partial s}(S(t),t).$$

We end up with an expression that does not depend on the function $\mu(S(t),t)$ at all. More important is that we have come to the conclusion that if we construct the portfolio like this, the differential $dV(t)$ does not depend on the stochastic $dS(t)$ part at all. We interpret this as the portfolio being riskless in a short time interval and thus, from our no-arbitrage assumption, the instantaneous return must be the risk-free rate r, and

$$dV(t) = rV(t)dt. \tag{11}$$

From our expression

$$V(t) = Y(t) - A(t)S(t) + B(t)e^{rt},$$

we get by multiplying through with r,

$$B(t)re^{rt} = r\left(V(t) - Y(t) + A(t)S(t)\right).$$

We know what $Y(t)$ and $A(t)$ is, and we get

$$B(t)re^{rt} = r\left(V(t) - f(S(t),t) + S(t)\frac{\partial f}{\partial s}(S(t),t)\right). \tag{12}$$

Now we replace $dV(t)$ by $rV(t)dt$ and insert (12) in (10), ending up with

$$rV(t)dt = \left(\frac{\partial f}{\partial t}(S(t),t) + \frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t) + r\left(V(t) - f(S(t),t)\right.\right.$$
$$\left.\left. + S(t)\frac{\partial f}{\partial s}(S(t),t)\right)\right) dt.$$

Here we have dt on both sides and a term $rV(t)$ on both sides, thus we end up with

$$0 = \frac{\partial f}{\partial t}(S(t),t) + \frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t)$$
$$-rS(t)f(S(t),t) + r\frac{\partial f}{\partial s}(S(t),t),$$

and we rearrange it so that we have it on the form it is usually written on,

$$\frac{\partial f}{\partial t}(S(t),t) = rf(S(t),t) - rS(t)\frac{\partial f}{\partial s}(S(t),t)$$
$$-\frac{1}{2}\sigma(S(t),t)^2 S(t)^2 \frac{\partial^2 f}{\partial s^2}(S(t),t). \tag{13}$$

Now, this will hold if f satisfies the partial differential equation

$$\frac{\partial f}{\partial t}(s,t) = rf(s,t) - rs\frac{\partial f}{\partial s}(s,t) - \frac{1}{2}\sigma(s,t)^2 s^2 \frac{\partial^2 f}{\partial s^2}(s,t), \tag{14}$$

with the boundary condition

$$f(s,T) = \max(s-K,0).$$

The difference from the usual BSM partial differential equation (5) is that we here have a function $\sigma(s,t)$ which depends on the underlying stock price and the time rather than being assumed constant.

Remark. We said that the portfolio was riskless in a short time interval, and that the instantaneous return must be the risk-free rate of return. There is an important thing to note about that. It means that a particular portfolio is only riskless at a point in time, and thus one has to rebalance the portfolio at all times to maintain this. It is an assumption made that this is possible.

5 Solving the partial differential equation for a given function σ

Section 4 will be based on [11], written by Gerald Recktenwald and [10], written by Morton and Mayers.

We now have a partial differential equation that determines the option price when the volatility function is known. On the contrary to when the volatility is assumed to be constant, this differential equation does not have an explicit solution. The aim of this section is to describe one way to numerically solve this partial differential equation given a volatility function. In this paper we use a Crank-Nicolson finite difference method.

5.1 Taylor's Theorem

The theorem and the idea of the proof originate from [12].

An important part in this section will be the use of Taylor's Theorem. This theorem shows that a function f can be approximated by a polynomial of degree n, and that we have a way to estimate the error, at least if we know bounds on $|f^{(n+1)}(x)|$.

Theorem 2. *Let $n \geq 1$ be an integer and suppose that f is a real function on [a,b], that $f^{(n)}$ is continuous on the interval [a,b] and that $f^{(n+1)}$ exists on (a,b). Let α and β be distinct points in [a,b] such that $\alpha < \beta$. We define the n-th order Taylor polynomial of the function at the point α as*

$$P(x) = \sum_{k=0}^{n} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k. \tag{15}$$

Then there exist a point $\xi \in (\alpha, \beta)$ such that

$$f(\beta) = P(\beta) + \frac{f^{(n+1)}(\xi)}{(n+1)!}(\beta-\alpha)^{n+1}. \tag{16}$$

Proof. Let M be defined as the number that satisfies

$$f(\beta) = P(\beta) + M(\beta-\alpha)^{n+1}$$

and put

$$g(x) = f(x) - P(x) - M(x-\alpha)^{n+1} \qquad (a \leq x \leq b). \tag{17}$$

If we can show that $(n+1)!M = f^{(n+1)}(\xi)$ for some $\xi \in (a,b)$ we are done. By (15) we have,

$$g^{(n+1)}(x) = f^{(n+1)}(x) - (n+1)!M \qquad (a < x < b) \tag{18}$$

hence, if $g^{(n+1)}(\xi) = 0$ for some $\xi \in (a, b)$ we are done. If $k \leq n$ we have, by (15), that $P^{(k)}(\alpha) = f^{(k)}(\alpha)$. This combined with (17) gives us

$$g(\alpha) = g'(\alpha) = \cdots = g^{(n)}(\alpha) = 0. \tag{19}$$

From our choice of M we know that $g(\beta) = 0$, and thus the mean value theorem says that $g'(\xi_1) = 0$ for some choice of $\xi_1 \in (\alpha, \beta)$. Since $g'(\alpha) = 0$, again by the mean value theorem, applied to the interval $[\alpha, \xi_1]$ we have that $g''(\xi_2) = 0$ for some $\xi_2 \in (\alpha, \xi_1)$. After $n + 1$ steps we arrive at the conclusion that $g^{(n+1)}(\xi_{n+1}) = 0$ for some $\xi_{n+1} \in (\alpha, \xi_n)$, that is, for some $\xi_{n+1} \in (\alpha, \beta)$. □

5.2 The big $\mathcal{O}$ notation

We base the definitions here on the work of [7].

Since we are going to solve our partial differential equation numerically we need to have control, or at least be aware, of how big the error, that will inevitably come from approximations, is. A convenient way to do this is with the big $\mathcal{O}$ notation.

Definition. $f(x) = \mathcal{O}(g(x))$ *means* $\exists c \in \mathbb{R}\ \forall x\ (|f(x)| \leq c|g(x)|)$.

In this definition we assume that the functions f and g have the same domain and codomain.

We can interpret $\mathcal{O}(g)$ as denoting a set,

$$\mathcal{O}(g) = \{f \mid \exists c \in \mathbb{R}\ \forall x,\ |f(x)| \leq c|g(x)|\},$$

and then $f(x) = \mathcal{O}(g(x))$ means $f \in \mathcal{O}(g)$, in other words the *equality* is not an ordinary equality, the equality works in one way only since it denotes that it belongs to a set.

Generally, whether $f(x) = \mathcal{O}(g(x))$ or not might depend on what interval we are working with. For example, if $f(x) = \frac{1}{x}$ and $g(x) = x$ then $f(x) = \mathcal{O}(g(x))$ if x takes values greater than 1, but if we look at an interval $(0, 1]$ it will not be true since f is unbounded as $x \to 0$.

In order to solve that problem one can give *side conditions* to the variable given, that is, put some kind of restrictions on the set, for example that $x \geq x_0$ or $x \leq x_0$ for some fixed value x_0. We will use the notation that $f(x) = \mathcal{O}(g(x))\ \forall x \in \Omega$ where Ω is the set for which this is true, of course this set will have to be specified each time.

We here write down some rules of manipulations on the big $\mathcal{O}$ notation and prove one which will be used consequently. One important note here is that the set we are working with will be $\Omega = [0, 1]$, that is, x will be restricted to values at most 1. In case one wishes to look at greater values of x, that is, greater than 1., then the rules of manipulations will still be true except for 1, where we would instead have $0 \leq m \leq m'$.

1. $x^m = \mathcal{O}(x^{m'})$ when $0 \leq m' \leq m$;
2. $\mathcal{O}(g(x)) + \mathcal{O}(f(x)) = \mathcal{O}(|g(x)| + |f(x)|)$;
3. $f(x) = \mathcal{O}(f(x))$;
4. $c\,\mathcal{O}(f(x)) = \mathcal{O}(f(x))$, if c is a constant;
5. $\mathcal{O}(\mathcal{O}(f(x))) = \mathcal{O}(f(x))$;
6. $\mathcal{O}(f(x))\mathcal{O}(g(x)) = \mathcal{O}(f(x)g(x))$;
7. $\mathcal{O}(f(x)g(x)) = f(x)\mathcal{O}(g(x))$.

1. is clear since $x^m \leq x^{m'}$ when $x \in [0, 1]$. We will here prove 2., and refer the interested reader to [7] for the rest of them.

Proof. Let $a(x) \in \mathcal{O}(g(x))$ and let $b(x) \in \mathcal{O}(f(x))$. Then by definition, for every $a(x)$ $\exists c_0$ such that $\forall x \in [0, 1]$, $|a(x)| \leq c_0|g(x)|$. Similarly $\exists d_0$ such that $|b(x)| \leq d_0|f(x)|$ $\forall x \in [0, 1]$.

Let $e_0 = \max(c_0, d_0)$. Every function on the left hand side of 2. is on the form $a(x) + b(x)$, and $|a(x) + b(x)| \leq |a(x)| + |b(x)| \leq c_0|g(x)| + d_0|f(x)| \leq e_0(|g(x)| + |f(x)|$, i.e. it is a member of the right hand side. □

We are of course going to use these results in some way. We will later use several, but finitely many, approximations, one at each point where we do an approximation. The total error will be the sum of these errors, and by result 2, that we proved, that error is easily described.

We will have error terms that are $\mathcal{O}(\Delta s)$, where Δs is the variable that we can control, and therefore we can choose to have it on the interval [0,1]. So, if we sum such terms we have, from 2, that $\mathcal{O}(\Delta s) + \mathcal{O}(\Delta s) = \mathcal{O}(\Delta s)$. From 1 and 2 we get that $\mathcal{O}((\Delta s)^k) + \mathcal{O}((\Delta s)^{k-1}) = \mathcal{O}((\Delta s)^{k-1})$ for $k > 1$. This is really all we need here, we now move on to the method.

5.3 Finite Difference Method

We begin with explaining what a finite difference method is.

Solving a differential equation with finite differences means that we replace the partial derivatives with a discrete approximation [11]. *Discrete* has the meaning that instead of $\frac{\partial f}{\partial s}$ we instead look at an approximation

$$\frac{\partial f}{\partial s} \approx \frac{f_{k+1} - f_k}{\Delta s}.$$

This means that we only evaluate the solution at a finite number of points, with a specified distance Δs from each other. f_k denotes an approximated solution that has been obtained using a finite difference method. The approximations obtained by the finite difference method of $f(s,t)$ at a point (s_k, t_j) will be denoted as f_k^j

The differential equation (14) has derivatives with respect to the time t and the stock s, as well as a second derivative with respect to the stock s. We assume that $f(s,t)$ is twice differentiable with respect to t, four times differentiable with respect to s and that the derivatives are bounded on the relevant interval.

We use N different points for the stock prices that are uniformly spaced on the interval $0 \leq s \leq 2s_{\text{init}}$. That is,

$$s_k := k\Delta s, \;\; k = 0, \cdots, N-1$$

$$\text{and } \Delta s := \frac{2s_{\text{init}}}{N-1}.$$

Similarly we have $0 \leq t \leq T$ and space it uniformly with

$$t_j := j\Delta t, \;\; j = 0, \cdots, M-1$$

$$\text{and } \Delta t := \frac{T}{M-1}.$$

Remark. Since we in section 5.2 used the interval [0,1] for our results, we will assume that from this point, Δs, $\Delta t \in [0,1]$. It could be so that, due to computer limits, this is not possible. That is, a computer might not be able to handle enough points in order to allow the distance between the points to be ≤ 1.

The changes that would come from having a different interval would not be great. The reason is that Δs is variable only in the sense that we can choose how many points we want to have. Thus, if the distance is not less

than or equal to 1, then it is greater than 1. This means that at every time we would like to calculate the error, Δs (or Δt) would be fixed. We could thus handle values of $\Delta s > 1$ separately. We would have to change some of the rules for the big $\mathcal{O}$ notation, for example 1. would be reversed. We have not done this here, but one can look in for example [7].

5.4 Approximating with Taylor's Theorem, forward difference

We will look at this only for one variable, s, but the result holds for both variables, it is only to simplify notation. Due to that we use the notation f(s) for this section. Furthermore, the notation of approximations will be written as, for example, $f_k - f_{k-1}$. That means that we really evaluate it at $f_k^j - f_{k-1}^j$, that is, we look at the change in one variable, and keep the other one fixed.

With Taylor's theorem we get,

$$f(s_k + \delta s) = f(s_k) + \delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(\xi)$$

for some $\xi \in (s_k, s_k + \delta s)$. δs is the change in s from s_k. This change δs can be chosen as one wishes. We want it to be equal to the distance between our points and thus let $\delta s = \Delta s$, making the change as big as the distance between our chosen points. Then we have that $s_k + \Delta s = s_{k+1}$. If we do that we get, for some $\xi \in (s_k, s_{k+1})$,

$$f(s_{k+1}) = f(s_k) + \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(\xi). \tag{20}$$

As we will show later, in the end of 5.8, we have access to initial conditions at the time T, that is, we know the value of the option at the expiration day.

As we are interested in starting at the end in time, we are really interested in going *back* from $f(s_{k+1})$ to $f(s_k)$, we thus rearrange (20),

$$f(s_k) = f(s_{k+1}) - \Delta s \frac{\partial f}{\partial s}(s_k) - \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(\xi). \tag{21}$$

The error of our partial derivative, that is, the error of going from $f(s_{k+1})$ to $f(s_k)$, is expressed by the error term with the unknown ξ. How big is this? Well, we do not know the exact value of $\frac{\partial^2 f}{\partial s^2}(\xi)$ but it does not really matter. It is bounded at least, so we can use the big $\mathcal{O}$ notation to express

the error,

$$f(s_k) = f(s_{k+1}) - \Delta s \frac{\partial f}{\partial s}(s_k) + \mathcal{O}((\Delta s)^2). \tag{22}$$

That is, the error of going from $f(s_{k+1})$ to $f(s_k)$, following the tangent at that point, is $\mathcal{O}((\Delta s)^2)$.

We could look at what happens if we take it one step further, that is, does the error increase as we step our way back?

$$f(s_{k-1}) = f(s_k) - \Delta s \frac{\partial f}{\partial s}(s_{k-1}) + \mathcal{O}((\Delta s)^2).$$

Here we can replace $f(s_k)$ by its corresponding approximation (22). We get,

$$f(s_{k-1}) = f(s_{k+1}) - \Delta s \frac{\partial f}{\partial s}(s_k) + \mathcal{O}((\Delta s)^2) - \Delta s \frac{\partial f}{\partial s}(s_{k-1}) + \mathcal{O}((\Delta s)^2)$$

$$= f(s_{k+1}) - \Delta s \frac{\partial f}{\partial s}(s_k) - \Delta s \frac{\partial f}{\partial s}(s_{k-1}) + \mathcal{O}((\Delta s)^2. \tag{23}$$

This procedure can be repeated as many times as needed. That is, the error might increase in absolute values, but in the big $\mathcal{O}$ notation it stays of the same magnitude regardless of how many steps we take.

This is however only part of the error that we face. It is important though. But we are also replacing the derivative with the slope between two points. We can solve (22) for the derivative, to see how big the error is of replacing it with the slope between two points,

$$\frac{\partial f}{\partial s}(s_k) = \frac{f(s_{k+1}) - f(s_k)}{\Delta s} + \mathcal{O}(\Delta s).$$

That is, the error of the replacement is of the magnitude

$$\frac{\partial f}{\partial s}(s_k) - \frac{f(s_{k+1}) - f(s_k)}{\Delta s} = \mathcal{O}(\Delta s).$$

We have just one more thing to deal with before we can leave this. We are not actually approximating the derivative with the true function values at the points s_k and s_{k+1} but with the approximated function values f_{k+1}, f_k. Does this have any impact on the error? As we saw in (23), the magnitude of the error did not depend on how far we had moved, that is, it did not depend on what the value of k was. Thus, assuming that we have access to one true initial value on the derivative, the error of the derivative will be,

$$\frac{\partial f}{\partial s}(s_k) = \frac{f_{k+1} - f_k + \mathcal{O}((\Delta s)^2)}{\Delta s} + \mathcal{O}(\Delta s),$$

or

$$\frac{\partial f}{\partial s}(s_k) - \frac{f_{k+1} - f_k}{\Delta s} = \mathcal{O}(\Delta s). \tag{24}$$

The error on the right hand side is called the *truncation error* of this finite difference approximation. We cannot get much further here by just focusing on the forward approximation alone. We need a differential equation to work with in order to actually work with the errors.

5.5 Approximating with Taylor's Theorem, backward difference

Instead of looking at Taylor's theorem where you move Δs forward we could put $\delta s = -\Delta s$ and would get the equation

$$f(s_{k-1}) = f(s_k) - \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(\xi)$$

for some $\xi \in [s_{k-1}, s_k]$. Here we have a similar situation to the forward difference. We will not go through the details again. But solving here for $\frac{\partial f}{\partial s}(s_k)$, replacing the function values with our approximated values $f_{k-1},\ f_k$ and using the big $\mathcal{O}$ notation to sum up all errors, we get,

$$\frac{\partial f}{\partial s}(s_k) = \frac{f_k - f_{k-1}}{\Delta s} + \mathcal{O}(\Delta s). \tag{25}$$

Again, the truncation error is $\mathcal{O}(\Delta s)$.

5.6 Approximating with Taylor's Theorem, central difference

If we combine two Taylor approximations we can get a truncation error which approaches 0 faster for certain intervals on Δs, namely the one we assume we have here, with $\Delta s \in [0, 1]$. We have to assume that $f^{(3)}$ exists now, and will now use Taylor's theorem to state two approximations around a point s_k for the second order Taylor polynomials,

$$f(s_{k+1}) = f(s_k) + \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(s_k) + \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(\xi) \tag{26}$$

for some $\xi \in [s_k, s_{k+1}]$, and

$$f(s_{k-1}) = f(s_k) - \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(s_k) - \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(\zeta) \tag{27}$$

for some $\zeta \in [s_{k-1}, s_k]$.

If we subtract equation (27) from equation (26) we get

$$f(s_{k+1}) - f(s_{k-1}) = 2\Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(\xi) + \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(\zeta)$$

and solving for $\frac{\partial f}{\partial s}(s_k)$ yields

$$\frac{\partial f}{\partial s}(s_k) = \frac{f(s_{k+1}) - f(s_{k-1})}{2\Delta s} - \frac{(\Delta s)^2}{3!2}\left(\frac{\partial^3 f}{\partial s^3}(\xi) + \frac{\partial^3 f}{\partial s^3}(\zeta)\right).$$

We proceed as earlier, replacing the error terms with big $\mathcal{O}$ notation and replacing the function values with approximated solutions obtained with the finite difference method and we get,

$$\frac{\partial f}{\partial s}(s_k) = \frac{1}{2}\frac{f_{k+1} - f_{k-1}}{\Delta s} + \mathcal{O}((\Delta s)^2). \tag{28}$$

The truncation error is here $\mathcal{O}((\Delta s)^2)$.

5.7 Central difference, second order

In order to get a good approximation of the second order derivative we have to assume that $f^{(4)}$ exists, we can then write down two third order Taylor polynomials around a point s_k with Taylor's theorem,

$$f(s_{k+1}) = f(s_k) + \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(s_k) + \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(s_k) + \frac{(\Delta s)^4}{4!}\frac{\partial^4 f}{\partial s^4}(\xi)$$

for some $\xi \in [s_k, s_{k+1}]$, and

$$f(s_{k-1}) = f(s_k) - \Delta s \frac{\partial f}{\partial s}(s_k) + \frac{(\Delta s)^2}{2!}\frac{\partial^2 f}{\partial s^2}(s_k) - \frac{(\Delta s)^3}{3!}\frac{\partial^3 f}{\partial s^3}(s_k) + \frac{(\Delta s)^4}{4!}\frac{\partial^4 f}{\partial s^4}(\zeta)$$

for some $\zeta \in [s_{k-1}, s_k]$.

Adding these two expressions yields,

$$f(s_{k+1}) + f(s_{k-1}) = 2f(s_k) + (\Delta s)^2\frac{\partial^2 f}{\partial s^2}(s_k) + \frac{(\Delta s)^4}{4!}\frac{\partial^4 f}{\partial s^4}(\xi) + \frac{(\Delta s)^4}{4!}\frac{\partial^4 f}{\partial s^4}(\zeta).$$

Solving for $\frac{\partial^2 f}{\partial s^2}$ and proceeding as earlier with big $\mathcal{O}$ notations and substituting for approximate solutions we get

$$\frac{\partial^2 f}{\partial s^2}(s_k) = \frac{f_{k+1} - 2f_k + f_{k-1}}{(\Delta s)^2} + \mathcal{O}((\Delta s)^2). \tag{29}$$

5.8 Crank-Nicolson Finite Difference Method

We now have approximations for all differentials in the partial differential equation. The Crank-Nicolson finite difference method uses the forward difference, equation (24), for $\frac{\partial f}{\partial t}$. For $\frac{\partial f}{\partial s}, \frac{\partial^2 f}{\partial s^2}$ an average of the central difference equations, (28) and (29), are used, evaluated at the current and the forward time step. This introduces an extra factor $\frac{1}{2}$ on both approximations.

We also evaluate $f(s,t)$ as $\frac{1}{2}(f_k^j + f_k^{j+1})$. The reason we want to use the average on the central differences is that the truncation errors will be $\mathcal{O}(\Delta t^2)$ and $\mathcal{O}(\Delta s^2)$. It goes beyond the scope of this paper to show that we indeed do have these truncation errors, but we refer the interested reader to [10].

The reason to use the current and future time steps is that we have at hand an end condition in time for the option price and will work our way backwards.

We replace the continuous derivatives in (14) with these finite difference approximations and end up with

$$\begin{aligned}\frac{f_k^{j+1} - f_k^j}{\Delta t} = \frac{1}{2}(rf_k^j + rf_k^{j+1}) - rs_k\frac{1}{4}\left[\frac{f_{k+1}^j - f_{k-1}^j + f_{k+1}^{j+1} - f_{k-1}^{j+1}}{\Delta s}\right] \\ - \frac{1}{4}\left[\frac{f_{k+1}^j - 2f_k^j + f_{k-1}^j + f_{k+1}^{j+1} - 2f_k^{j+1} + f_{k-1}^{j+1}}{(\Delta s)^2}\right]\sigma(s_k,t_j)^2 s_k^2,\end{aligned} \tag{30}$$

with the boundary condition

$$f(s,T) = \max(s-K, 0).$$

We rearrange equation (30) and sum up the coefficients to each f_k^j. In order for the equation to be easier to read we write σ_k^j instead of $\sigma(s_k,t_j)$, the function is still evaluated at those points though, it is not an approximated solution of the function. We end up with

$$
\begin{aligned}
f_{k-1}^j\left(\frac{(\sigma_k^j)^2 s_k^2}{4(\Delta s)^2}-\frac{rs_k}{4\Delta s}\right)+f_k^j\left(-\frac{1}{\Delta t}-\frac{(\sigma_k^j)^2 s_k^2}{2(\Delta s)^2}-\frac{1}{2}r\right) \\
+f_{k+1}^j\left(\frac{rs_k}{4\Delta s}+\frac{(\sigma_k^j)^2 s_k^2}{4(\Delta s)^2}\right) \\
= f_{k-1}^{j+1}\left(-\frac{(\sigma_k^j)^2 s_k^2}{4(\Delta s)^2}+\frac{rs_k}{4\Delta s}\right)+f_k^{j+1}\left(-\frac{1}{\Delta t}+\frac{(\sigma_k^j)^2 s_k^2}{2(\Delta s)^2}+\frac{1}{2}r\right) \\
+f_{k+1}^{j+1}\left(-\frac{rs_k}{4\Delta s}-\frac{(\sigma_k^j)^2 s_k^2}{4(\Delta s)^2}\right).
\end{aligned} \tag{31}
$$

In section 5.3 we defined $s_k := k\Delta s$, $k = 0, 1, \cdots, N-1$. We replace s_k with $k\Delta s$ and multiply both sides by Δt to end up with

$$
\begin{aligned}
f_{k-1}^j\left(\frac{\Delta t}{4}(\sigma_k^j)^2 k^2-\frac{\Delta t}{4}rk\right)+f_k^j\left(-1-\frac{\Delta t}{2}(\sigma_k^j)^2 k^2-\frac{\Delta t}{2}r\right) \\
+f_{k+1}^j\left(\frac{\Delta t}{4}rk+\frac{\Delta t}{4}(\sigma_k^j)^2 k^2\right) \\
= f_{k-1}^{j+1}\left(-\frac{\Delta t}{4}(\sigma_k^j)^2 k^2+\frac{\Delta t}{4}rk\right)+f_k^{j+1}\left(-1+\frac{\Delta t}{2}(\sigma_k^j)^2 k^2+\frac{\Delta t}{2}r\right) \\
+f_{k+1}^{j+1}\left(-\frac{\Delta t}{4}rk-\frac{\Delta t}{2}(\sigma_k^j)^2 k^2\right).
\end{aligned} \tag{32}
$$

This can be written on a more compact form

$$
\begin{aligned}
a_k^j f_{k-1}^j+(-1-b_k^j)f_k^j+c_k^j f_{k+1}^j \\
= \quad -a_k^j f_{k-1}^{j+1}+(-1+b_k^j)f_k^{j+1}-c_k^j f_{k+1}^{j+1},
\end{aligned} \tag{33}
$$

where

$$
\begin{aligned}
a_k^j &= \frac{\Delta t}{4}((\sigma_k^j)^2 k^2-rk), \\
b_k^j &= \frac{\Delta t}{2}((\sigma_k^j)^2 k^2+r), \\
c_k^j &= \frac{\Delta t}{4}((\sigma_k^j)^2 k^2+rk).
\end{aligned}
$$

Note that this means that if one has a volatility function σ that depends on the time t, then a_k^j, b_k^j, c_k^j would differ for different j. We will later make the assumption that σ depends solely on s, so we will deal with them as being

independent of j, and that we therefore can write them as a_k, b_k, c_k.

We express the system of equations (33) as $Df^j = Ef^{j+1}$. This is a tridiagonal system

$$\begin{bmatrix} (-1-b_0) & c_0 & 0 & \dots & 0 & 0 \\ a_1 & (-1-b_1) & c_1 & \dots & 0 & 0 \\ 0 & a_2 & (-1-b_2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (-1-b_{N-2}) & c_{N-2} \\ 0 & 0 & 0 & \dots & a_{N-1} & (-1-b_{N-1}) \end{bmatrix} \underbrace{\begin{bmatrix} f_0^j \\ f_1^j \\ \vdots \\ f_{N-2}^j \\ f_{N-1}^j \end{bmatrix}}_{=f^j}$$

$$= \begin{bmatrix} (-1+b_0) & -c_0 & 0 & \dots & 0 & 0 \\ -a_1 & (-1+b_1) & -c_1 & \dots & 0 & 0 \\ 0 & -a_2 & (-1+b_2) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (-1+b_{N-2}) & -c_{N-2} \\ 0 & 0 & 0 & \dots & -a_{N-1} & (-1+b_{N-1}) \end{bmatrix} \underbrace{\begin{bmatrix} f_0^{j+1} \\ f_1^{j+1} \\ \vdots \\ f_{N-2}^{j+1} \\ f_{N-1}^{j+1} \end{bmatrix}}_{=f^{j+1}}$$

From our initial condition we know the values on the elements of the vector f^{M-1}, they are $\max(s_k - K, 0)$, $k = 0, \cdots, N-1$. Then we can iterate our way back from f^{M-1}. The elements of D and E can be calculated when we have access to a volatility function.

6 Nonlinear least squares problem

This section is based on chapter 17 in [14], written by Amol Sasane and Krister Svanberg.

As we will see in section 8, there will be a need for a method to solve a nonlinear least squares problem. This section will be devoted to explaining one method that can be used to solve such problems.

6.1 Defining the problem

We have a mathematical model describing an option price. We also have at hand measurement data of option prices, and our problem will be to fit our model to the measured data.

Assume that we have access to n different observed option prices. Each

corresponding to a distinct strike price K_j[1]. We define $\overline{v}_j := \overline{v}(K_j)$, $j = 1, \cdots, n$, where $\overline{v}(K_j)$ is the observed price of an option with strike price K_j.

The corresponding theoretical option value, given by solving the partial differential equation for a strike price K_j, will be denoted by $v_j(c(s,t;\overline{\sigma}))$. We will show what we mean with $c(s,t;\overline{\sigma})$ later on, in section 7. For now we settle with that c denotes a volatility function and that $\overline{\sigma}$ is a vector containing m elements, each element corresponding to a volatility value σ_i. The vector $\overline{\sigma}$ is the one we wish to calibrate in order to fit our model to the observed data.

We define $h_j(\overline{\sigma}) := \overline{v}_j - v_j(c(\overline{\sigma}))$, so h_j is the difference between the observed price and the price predicted by our model for a strike price K_j. In a perfect world we would like to have

$$\begin{cases} h_1(\overline{\sigma}) = 0 \\ \vdots \\ h_n(\overline{\sigma}) = 0. \end{cases}$$

However we probably do not have a perfect model, and have to settle with an error. We cannot be sure to be able to match the observed values perfectly, and in this paper we even seek a volatility function with other properties than just matching the option prices. Instead we will try to minimize the squared differences and work with a nonlinear least squares problem

$$\text{minimize } F(\overline{\sigma}) := \frac{1}{2}\sum_{j=1}^{n}(h_j(\overline{\sigma}))^2. \tag{34}$$

It is a nonlinear problem because h_j are nonlinear functions of $\overline{\sigma}$.

6.2 Gauss-Newton

The method we are going to use is Gauss-Newton, which is an iterative method, and therefore it will suffice to show how we go from one iterative point $\overline{\sigma}^{(k)}$ to $\overline{\sigma}^{(k+1)}$.

At an iteration point $\overline{\sigma}^{(k)}$ we approximate every function h_i with its first order Taylor polynomial at $\overline{\sigma}^{(k)}$,

$$h_j(\overline{\sigma}) \approx h_j(\overline{\sigma}^{(k)}) + \nabla h_j(\overline{\sigma}^{(k)})(\overline{\sigma} - \overline{\sigma}^{(k)}),\ \ j = 1, \cdots, n.$$

Let $d := \overline{\sigma} - \overline{\sigma}^{(k)}$, then the Taylor approximation can be written as

$$h_j(\overline{\sigma}^{(k)} + d) \approx h_j(\overline{\sigma}^{(k)}) + \nabla h_j(\overline{\sigma}^{(k)})d,\ \ j = 1, \cdots, n.$$

[1]It does not have to be distinct strike prices, they could differ in time to maturity as well, thus one could have different observed option prices for the same strike price K.

In order to write this in a more compact form, we define

$$h(\overline{\sigma}) = \begin{bmatrix} h_1(\overline{\sigma}) \\ \vdots \\ h_n(\overline{\sigma}) \end{bmatrix} \quad \text{and} \quad \nabla h(\overline{\sigma}) = \begin{bmatrix} \nabla h_1(\overline{\sigma}) \\ \vdots \\ \nabla h_n(\overline{\sigma}) \end{bmatrix} = \begin{bmatrix} \frac{\partial h_1}{\partial \overline{\sigma}_1} & \cdots & \frac{\partial h_1}{\partial \overline{\sigma}_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial \overline{\sigma}_1} & \cdots & \frac{\partial h_n}{\partial \overline{\sigma}_m} \end{bmatrix}.$$

We are now able to write the minimizing problem as

$$\text{minimize } F(\overline{\sigma}) = \frac{1}{2}||h(\overline{\sigma})||^2, \tag{35}$$

and our first order polynomial approximation is

$$h(\overline{\sigma}^{(k)} + d) \approx h(\overline{\sigma}^{(k)}) + \nabla h(\overline{\sigma}^{(k)})d.$$

Using this approximation in the objective function F we get

$$\begin{aligned} F(\overline{\sigma}^{(k)} + d) &= \frac{1}{2}||h(\overline{\sigma}^{(k)} + d)||^2 \\ &\approx \frac{1}{2}||h(\overline{\sigma}^{(k)}) + \nabla h(\overline{\sigma}^{(k)})d||^2 \\ &= \frac{1}{2}||A^{(k)}d - b||^2, \end{aligned} \tag{36}$$

where $A^{(k)} := \nabla h(\overline{\sigma}^{(k)})$ and $b := -h(\overline{\sigma}^{(k)})$.

In the Gauss-Newton method one wishes to minimize the right side of (36) for the vector $d \in \mathbb{R}^m$:

$$\text{minimize } \frac{1}{2}||A^{(k)}d - b||^2. \tag{37}$$

But this means we want to solve a linear least square problem, the solution to this can be found in for example [14] and is given by the normal equation $(A^{(k)})^T A^{(k)} d = (A^{(k)})^T b$. Replacing $A^{(k)}$ and b with the corresponding expressions of $h(\overline{\sigma})$ yields

$$(\nabla h(\overline{\sigma}^{(k)}))^T \nabla h(\overline{\sigma}^{(k)})d = -(\nabla h(\overline{\sigma}^{(k)}))^T h(\overline{\sigma}^{(k)}). \tag{38}$$

Let a solution of this normal equation be called $d^{(k)}$, then the next iteration point is

$$\overline{\sigma}^{(k+1)} = \overline{\sigma}^{(k)} + d^{(k)}. \tag{39}$$

7 The volatility function

We have earlier said that in order to solve the partial differential equation numerically, we need a volatility function. In the previous section we mentioned that the function h would depend on a vector $\overline{\sigma}$, and that this vector is going to be updated when we solve the minimizing problem that is to fit our model to data from the market.

This section will be devoted to show what kind of function we want this volatility function to be, and thus what kind of function we will recreate. We will do this by solving a minimizing problem.

At this point we make a quite big simplification that makes the problem easier to solve. The effect this has on accuracy of the solution has not been looked at. The assumption made is that the volatility function, σ, does not depend on time, i.e. it is a function of only one variable, the underlying stock.

Remark. In case someone with a background in financial economics read this we do not claim that the volatility is best, or even well, described as a function of solely the stock price. It is a pure simplification made to enable this paper to be written within the time frame. More than that, the method given does not really depend on *what* one assumes that the volatility depend on, the simplification made is really that it depends on only one variable.

7.1 Some basic theorems

We start by stating some theorems that will be needed when deciding what kind of function it is we seek. Both theorems originate from [15]. We only prove theorem 3 here since we have changed the statement a bit from the original one, the idea of the proof is the same though.

Theorem 3. *Suppose that x_1 and x_2 are fixed constants with $x_1 < x_2$, $G(x)$ is a continuous function on $[x_1, x_2]$ such that for every continuous and differentiable function η, with the property $\eta(x_1) = \eta(x_2) = \eta'(x_1) = \eta'(x_2) = 0$, we have*

$$\int_{x_1}^{x_2} G(x)\eta(x)dx = 0. \tag{40}$$

Then

$$G(x) = 0 \ \forall x \in [x_1, x_2]. \tag{41}$$

Proof. If we can find one function $\eta(x)$ for which (40) is false when (41) does not hold, we are done. Let us therefore suppose that (41) does not hold, that is, there exist a particular value $x' \in (x_1, x_2)$ such that $G(x')$ differs

from 0 and (40) is still true.

For the sake of definiteness, suppose $G(x') > 0$. G is a continuous function, therefore there exists an interval around x', say $[x_1', x_2']$, $x_1 \leq x_1' < x' < x_2' \leq x_2$, such that $G(x) > 0 \; \forall x \in [x_1', x_2']$. We can now consider the function η defined by

$$\eta(x) = \begin{cases} 0 & \text{for } x_1 \leq x \leq x_1', \\ (x - x_1')^2(x - x_2')^2 & \text{for } x_1' \leq x \leq x_2', \\ 0 & \text{for } x_2' \leq x \leq x_2. \end{cases} \tag{42}$$

This particular function satisfy the properties demanded of η, and the integral (40) becomes

$$\int_{x_1'}^{x_2'} G(x)(x - x_1')^2(x - x_2')^2 dx. \tag{43}$$

Since $G(x) > 0 \; \forall x \in [x_1', x_2']$, this integral is clearly positive, and thus violate (40). We can reach a similar contradiction if $G(x) < 0$ on the interval, and the proof is complete. □

Theorem 4. *Suppose that $\partial f/\partial \epsilon$ is a continuous function of ϵ and of $x \in [x_1, x_2]$. Suppose also that x_1, x_2 are differentiable with respect to ϵ. If*

$$I = I(\epsilon) = \int_{x_1(\epsilon)}^{x_2(\epsilon)} f(x, \epsilon) dx,$$

then

$$\frac{dI}{d\epsilon} = I'(\epsilon) = f(x_2, \epsilon)\frac{dx_2}{d\epsilon} - f(x_1, \epsilon)\frac{dx_1}{d\epsilon} + \int_{x_1(\epsilon)}^{x_2(\epsilon)} \frac{\partial f}{\partial \epsilon}(x, \epsilon) dx.$$

If we have an integral where x_1, x_2 are independent of ϵ, the derivative is reduced to the integral term.

7.2 Defining the problem

The main idea of what kind of function we are trying to recreate comes from [3], a work of Thomas F. Coleman, Yuying Li and Arun Verma.

We seek to reconstruct the volatility function. To do that, we need to have some restrictions on what kind of function we seek. Here we want to get a fairly smooth function and therefore use splines which have good

approximation properties. In order to get one we follow ideas from physics and seek a volatility function $\sigma \in P$ that minimizes

$$\frac{1}{2}\int_0^{s_{m+1}} (\sigma''(s))^2 ds, \text{ subject to } \sigma(s_i) = \sigma_i, \ i = 1, \cdots, m$$

where σ_i are observed volatility values and $s_{m+1} > s_m$ is a fixed number in $\mathbb{R}$. P is a functional room consisting of functions that are $C^2[0, s_{m+1}]$. Apart from belonging to P we also demand that all functions σ has the property $\sigma(s_i) = \sigma_i$, $i = 1, ..., m$. As we noted in section 6, the vector $\overline{\sigma}$ denotes a vector with σ_i as the elements. This vector is what will be variable in the least squares problem defined in sec 6 that we are going to give a scheme to solve later.

We make a definition of the kind of function that will later be proved to be the minimizer of this integral.

Definition. A function $\sigma \in C^2[a, b]$ is called a *cubic spline* on $[a, b]$ if $[a, b] = [c_0, c_1] \cup [c_1, c_2] \cup \cdots \cup [c_{m-1}, c_m]$ and if σ consists of cubic polynomials p_i on each subinterval $[c_i, c_{i+1}]$, $i = 0, 1, ..., m - 1$.

If the polynomial $\sigma(s)$ has the property that $\sigma(s_i) = \sigma_i$ for given values σ_i, then σ is called an *interpolating cubic spline.*

If furthermore the interpolating cubic spline has the property $p_0''(c_0) = p_{m-1}''(c_m) = 0$ then it is said to have a *natural spline end condition.*

7.3 Finding the minimizers behaviour at (s_i, s_{i+1})

Here we use methods presented in [15], chapter 3, when trying to find properties of the minimizer.

Let us assume that the function $\sigma_* \in P$ is the actual minimizing function and introduce a one-parameter family of comparison functions,

$$\sigma(s) = \sigma_*(s) + \epsilon h(s),$$

where ϵ is the parameter of the family. We want σ to be a candidate to our volatility function $\forall \epsilon$. This can be achieved if $h(s_i) = 0$, $i = 1, ..., m$ and $h(s) \in C^2[0, s_{m+1}]$. These are the only restrictions we need to have on h, and apart from that it can be arbitrarily chosen. We denote the set with all functions having these properties as P'. This choice of $h \in P'$ gives us $\sigma(s) \in C^2[0, s_{m+1}]$ and thus $\sigma \in P$. Also $\sigma(s_i) = \sigma_*(s_i) = \sigma_i$, $i = 1, \cdots, m$, and so σ is a candidate for our minimizing function $\forall \epsilon$.

Now we replace σ_* by σ in the integrand and we get a comparison integral

$$I(\epsilon) = \frac{1}{2}\int_0^{s_{m+1}} \left(\sigma_*''(s) + \epsilon h''(s)\right)^2 ds.$$

Now this integral is clearly a function of one variable, ϵ, and since we have assumed σ_* to be the minimizing function, we see that the minimum of $I(\epsilon)$ must be when $\epsilon = 0$. Thus we get that

$$I'(0) = 0. \tag{44}$$

With Theorem 4 we can calculate the derivative with respect to ϵ,

$$\begin{aligned} I'(\epsilon) &= \frac{1}{2}\int_0^{s_{m+1}} \left(\frac{\partial}{\partial\epsilon}\left((\sigma_*''(s) + \epsilon h''(s)\right)^2)\right) ds \\ &= \frac{1}{2}\int_0^{s_{m+1}} \left(2(\sigma_*''(s) + \epsilon h''(s))\frac{\partial}{\partial\epsilon}(\sigma_*''(s) + \epsilon h''(s))\right) ds \\ &= \int_0^{s_{m+1}} \left((\sigma_*''(s) + \epsilon h''(s))h''(s)\right) ds. \end{aligned} \tag{45}$$

We have used the fact that

$$\sigma''(s) = \sigma_*''(s) + \epsilon h''(s).$$

(45) combined with (44) gives us

$$0 = \int_0^{s_{m+1}} \sigma_*''(s)h''(s)ds.$$

We start by focusing on some interval $[s_i, s_{i+1}]$, $1 \le i \le m-1$. Since this should hold for all h with the restrictions we have imposed, we can in particular choose h such that $h(s) = 0$ outside of this interval, and we can also demand that $h(s_i) = h(s_{i+1}) = h'(s_i) = h'(s_{i+1}) = 0$.

Assume now that $\sigma_* \in C^4$ on every subinterval, we can proceed with integration by parts and derive information about the possible optimizer σ_*.

$$\begin{aligned} 0 &= \int_{s_i}^{s_{i+1}} \sigma_*''(s)h''(s)ds \\ &= \underbrace{[h'(s)\sigma_*''(s)]_{s_i}^{s_{i+1}}}_{=0} - \int_{s_i}^{s_{i+1}} \sigma_*^{(3)}(s)h'(s)ds. \end{aligned}$$

The first term is equal to 0 since $h'(s_i) = h'(s_{i+1}) = 0$. Can we use theorem 3 and conclude that $\sigma_*^{(3)} = 0$ on (s_i, s_{i+1})? $h'(s)$ has the property $h'(s_i) =$

$h'(s_{i+1}) = 0$, but it has one extra restriction, namely that the integral $\int_{s_i}^{s_{i+1}} h'(s)ds = 0$ and thus we cannot use theorem 3. We integrate by parts once more and get

$$-\underbrace{[h(s)\sigma_*^{(3)}(s)]_{s_i}^{s_{i+1}}}_{=0} + \int_{s_i}^{s_{i+1}} \sigma_*^{(4)}(s)h(s)ds.$$

Here the first term vanishes since $h(s_i) = h(s_{i+1}) = 0$. We are left with

$$\int_{s_i}^{s_{i+1}} \sigma_*^{(4)}(s)h(s)ds = 0.$$

Both σ_* and h fulfils the requirements in Theorem 3 so we conclude that $\sigma_*^{(4)}(s) = 0$ for $s \in [s_i, s_{i+1}]$, and thus that it is a third degree polynomial on this subinterval. Since we did not choose the interval in any particular way this holds for all subintervals. We conclude that σ_* consists of third degree polynomials, that may differ from each other between each subinterval, but that have the same values at the connecting points.

At the intervals $[0, s_1], [s_m, s_{m+1}]$ we only have a single point to pass through. The function that minimizes the integral and passes through a single point must be a first degree polynomial. Thus, on both the first and the last interval, σ_* is a first degree polynomial.

We know more about σ_* though. Since we assumed it to be C^2, at each point s_i, $i = 1, \cdots, m$, we have that the possibly different third degree polynomials must have the same slope as well as the same slope on their derivatives.

Now we have a candidate for our minimizing function, is this then really the minimizing function?

7.4 Confirming the minimizer

The definition and theorem of this section originates from [13], where one can also find the proof of the theorem.

We wish to show that this minimizing function is a global minimizer, and that it thus is the function we seek.

Definition. Let X be a normed space. A function $I : X \to \mathbb{R}$ where,

for all $x_1, x_2 \in X$, and all $\alpha \in [0,1]$,

$$I(\alpha x_1 + (1-\alpha)x_2) \leq \alpha I(x_1) + (1-\alpha)I(x_2)$$

is said to be convex. If, for all $x_1, x_2 \in X$ with $x_1 \neq x_2$, and all $\alpha \in (0,1)$,

$$I(\alpha x_1 + (1-\alpha)x_2) < \alpha I(x_1) + (1-\alpha)I(x_2)$$

then I is said to be strictly convex.

Theorem 5. *Let X be a normed space, and let $I : X \to \mathbb{R}$ be differentiable. Suppose that I is convex. If $x_0 \in X$ is such that $I'(x_0) = 0$, then I has a global minimum at x_0. If furthermore I is strictly convex then the global minimizer is unique.*

We know that our function σ_* is a candidate to be the minimizer. If our integral is a strictly convex functional then we know that this function is a unique global minimizer.

We start by examining the function $g(x) = x^2$, $x \in \mathbb{R}$, since inside the integral we have $g(\sigma''(s))$. We know that $g''(x) = 2\ \forall x \in \mathbb{R}$, which is greater than 0 and thus g is strictly convex. From our definition of strictly convex, this means that for $\alpha \in (0,1)$, when $x_1 \neq x_2$, $g(\alpha x_1 + (1-\alpha)x_2) < \alpha g(x_1) + (1-\alpha)g(x_2)$.

Now, let $\tilde{\sigma}_1, \tilde{\sigma}_2 \in P$ be functions such that $\tilde{\sigma}_1 \neq \tilde{\sigma}_2$ and $\tilde{\sigma}_1(s_i) = \tilde{\sigma}_2(s_i) = \sigma_i$, $i = 1, \cdots, m$. Then they are any two functions which could be the minimizing function. Let

$$\Psi(\sigma) = \frac{1}{2}\int_0^{s_{m+1}} (\sigma''(s))^2 ds$$

and let $\alpha \in (0,1)$. Then,

$$\begin{aligned}
\Psi(\alpha\tilde{\sigma}_1 + (1-\alpha)\tilde{\sigma}_2) &= \frac{1}{2}\int_0^{s_{m+1}} (\alpha\tilde{\sigma}_1''(s) + (1-\alpha)\tilde{\sigma}_2''(s))^2 ds \\
&\leq \frac{1}{2}\int_0^{s_{m+1}} (\alpha(\tilde{\sigma}_1''(s))^2 + (1-\alpha)(\tilde{\sigma}_2''(s))^2) ds \\
&= \frac{1}{2}\int_0^{s_{m+1}} \alpha\tilde{\sigma}_1''(s)^2 ds + \frac{1}{2}\int_0^{s_{m+1}} (1-\alpha)\tilde{\sigma}_2''(s)^2 ds \\
&= \alpha\frac{1}{2}\int_0^{s_{m+1}} \tilde{\sigma}_1''(s)^2 ds + (1-\alpha)\frac{1}{2}\int_0^{s_{m+1}} \tilde{\sigma}_2''(s)^2 ds \\
&= \alpha\Psi(\tilde{\sigma}_1) + (1-\alpha)\Psi(\tilde{\sigma}_2).
\end{aligned}$$

When is the inequality a strict inequality? Since g is strictly convex, the inequality is a strict inequality unless $\tilde{\sigma}_1''(s) = \tilde{\sigma}_2''(s)\ \forall s \in [0, s_{m+1}]$. Let

us see then if this equality can hold for any two different σ that we are interested in.

$$\tilde{\sigma}_1''(s) = \tilde{\sigma}_2''(s) \quad \Leftrightarrow \quad \tilde{\sigma}_1'(s) = \tilde{\sigma}_2'(s) + c$$

for some constant c. If c = 0, then $\tilde{\sigma}_1(s) = \tilde{\sigma}_2(s) + d$, but $\tilde{\sigma}_1(s_i) = \tilde{\sigma}_2(s_i) = \sigma_i$, which implies that $d = 0$. Then $\tilde{\sigma}_1 = \tilde{\sigma}_2$, contrary to our choice of $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$.

If c is different from 0, assume for definiteness that $c > 0$, then we have $\tilde{\sigma}_1'(s) < \tilde{\sigma}_2'(s)$. This implies that if $\tilde{\sigma}_1(s_1) = \tilde{\sigma}_2(s_1) = \sigma_1$ then $\tilde{\sigma}_1(s_2) < \tilde{\sigma}_2(s_2)$, hence they cannot both be equal to σ_2, which is a contradiction to our choice of $\tilde{\sigma}_1, \tilde{\sigma}_2$. This is only true if there are more than one point we must pass through. An analogous argument hold if $c < 0$ and thus, if we have at least two points to take into consideration, our functional is strictly convex, and σ_* is indeed the unique global minimizer.

If we only do have one point we must pass through, the minimizer of the integral is any straight line that passes through that point since then the integral is 0.

7.5 Finding the coefficients of the polynomials

Now we have everything we need in order to find the coefficients of our polynomials. We know that the minimizing function σ_* consists of, possibly different, third degree polynomials on the intervals $[s_i, s_{i+1}]$, $i = 1, .., m-1$. We make a convenient definition.

$$p_i(s) := \sigma_*(s) \text{ for } s \in [s_i, s_{i+1}], \ i = 0, 1, ..., m$$

where $s_0 := 0$ (the reason it needs to be defined is that we do not actually have a point s_0).

What we have done is simply to let each subinterval be represented by a polynomial of its own. We have that $p_i(s) = A_i(s - s_i)^3 + B_i(s - s_1)^2 + C_i(s - s_i) + D_i$, $i = 0, ..., m$, where the coefficients may be different for different i. The reason to write $(s - s_i)$ instead of just s is that it becomes very easy to determine the value of D_i due to the fact that $\sigma_*(s_i) = \sigma_i$, $i = 1, \cdots, m$.

For p_0 and p_m we have that $p_0(s) = C_0 s + D_0$ and $p_m(s) = C_m(s - s_m) + D_m$ since we know that they are first degree polynomials.

Now we write up everything we know about the polynomials p_i.

$$\begin{cases} p_i(s_i) = \sigma_i, \\ p_i(s_i) = p_{i-1}(s_i) \\ p_i'(s_i) = p_{i-1}'(s_i) \\ p_i''(s_i) = p_{i-1}''(s_i) \end{cases} \quad \text{for } i = 1, ..., m. \tag{46}$$

The third and fourth condition follows since $\sigma_* \in C^2[0, s_{m+1}$. The fourth condition, together with the fact that p_0 and p_m are first degree polynomials, gives us that $p_1''(s_1) = p_{m-1}''(s_m) = 0$. This means that we have a natural interpolating cubic spline on the interval $[s_1, s_m]$.

We continue with calculating the derivatives which are the following,

$$\begin{cases} p_i(s) = A_i(s - s_i)^3 + B_i(s - s_i)^2 + C_i(s - s_i) + D_i \\ p_i'(s) = 3A_i(s - s_i)^2 + 2B_i(s - s_i) + C_i \\ p_i''(s) = 6A_i(s - s_i) + 2B_i \\ p_0(s) = C_0 s + D_0 \\ p_0'(s) = C_0 \\ p_m(s) = C_m(s - s_m) + D_m \\ p_m'(s) = C_m. \end{cases}$$

We note that $p_i(s_i) = D_i$, $p_i'(s_i) = c_i$ and $p_i''(s_i) = 2B_i$. Inserting this into (46) yields

$$\begin{cases} D_i = \sigma_i \\ D_i = A_{i-1}(s_i - s_{i-1})^3 + B_{i-1}(s_i - s_{i-1})^2 + C_{i-1}(s_i - s_{i-1}) + D_{i-1} \\ C_i = 3A_{i-1}(s_i - s_{i-1})^2 + 2B_{i-1}(s_i - s_{i-1}) + C_{i-1} \\ 2B_i = 6A_{i-1}(s_i - s_{i-1}) + 2B_{i-1}, \end{cases}$$

and we rewrite it slightly to end up with

$$\begin{cases} D_i = \sigma_i \\ A_{i-1}(s_i - s_{i-1})^3 + B_{i-1}(s_i - s_{i-1})^2 + C_{i-1}(s_i - s_{i-1}) = D_i - D_{i-1} \\ 3A_{i-1}(s_i - s_{i-1})^2 + 2B_{i-1}(s_i - s_{i-1}) + C_{i-1} - C_i = 0 \\ 6A_{i-1}(s_i - s_{i-1}) + 2B_{i-1} - 2B_i = 0. \end{cases} \tag{47}$$

We leave the coefficients C_0 and D_0 for now, and deal with them a little bit later. The rest of this system of equations can be explicitly solved for

A_{i-1}, B_{i-1} and C_{i-1} in terms of the coefficients A_i, B_i and C_i. We express the system of equations in a matrix where the first column corresponds to the coefficients of A_{i-1}, the second the coefficients of B_{i-1} and the third the coefficients of C_{i-1}. We denote the distance $s_i - s_{i-1}$ as Δs_i. The system is

$$\left(\begin{array}{ccc|c} (\Delta s_i)^3 & (\Delta s_i)^2 & \Delta s_i & \sigma_i - \sigma_{i-1} \\ 3(\Delta s_i)^2 & 2(\Delta s_i) & 1 & C_i \\ 6\Delta s_i & 2 & 0 & 2B_i \end{array}\right).$$

Let R_1, R_2 and R_3 denote row 1, 2 and 3. If we subtract $\Delta s_i R_3$ from R_2 and $\frac{1}{2}(\Delta s_i)^2 R_3$ from R_1 we get

$$\left(\begin{array}{ccc|c} -2(\Delta s_i)^3 & 0 & \Delta s_i & \sigma_i - \sigma_{i-1} - B_i(\Delta s_i)^2 \\ -3(\Delta s_i)^2 & 0 & 1 & C_i - 2B_i\Delta s_i \\ 6\Delta s_i & 2 & 0 & 2B_i \end{array}\right).$$

Now we divide R_3 by 2, and we subtract $-\Delta s_i R_2$ from R_1 and get

$$\left(\begin{array}{ccc|c} (\Delta s_i)^3 & 0 & 0 & \sigma_i - \sigma_{i-1} + B_i(\Delta s_i)^2 - C_i\Delta s_i \\ -3(\Delta s_i)^2 & 0 & 1 & C_i - 2B_i\Delta s_i \\ 3\Delta s_i & 1 & 0 & B_i \end{array}\right).$$

Finally, adding $\frac{3}{\Delta s_i} R_1$ to R_2 as well as subtracting $\frac{3}{(\Delta s_i)^2} R_1$ from R_3, and then dividing R_1 by $(\Delta s_i)^3$ yields

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{\sigma_i - \sigma_{i-1} + B_i(\Delta s_i)^2 - C_i\Delta s_i}{(\Delta s_i)^3} \\ 0 & 0 & 1 & -B_i\Delta s_i + \frac{\sigma_i - \sigma_{i-1}}{\Delta s_i} \\ 0 & 1 & 0 & -2B_i - 3\frac{\sigma_i - \sigma_{i-1} - C_i\Delta s_i}{(\Delta s_i)^2} \end{array}\right).$$

We have now showed how one determines every point. Only one question remains in order to say that the system has a unique solution, do we have a starting point? That is, do we have C_m and D_m? The question is no, not from the beginning. $D_m = \sigma_m$, but C_m is unknown, and thus all coefficients will be expressed in this term until we reach the end. Here we once again turn our attention to $p_0(s) = C_0 s + D_0$. From the system (47) we get,

$$\begin{cases} A_0(s_1)^3 + B_0(s_1)^2 + C_0(s_1) + D_0 = \sigma_1 \\ 3A_0(s_1)^2 + 2B_0(s_1) + C_0 = C_1 \\ 6A_0(s_1) + 2B_0 = 2B_1. \end{cases}$$

But $A_0 = B_0 = 0$, so this turns into

$$\begin{cases} C_0(s_1) + D_0 = \sigma_1 \\ C_0 = C_1 \\ 0 = 2B_1. \end{cases}$$

Since B_1 is unknown solely in C_m we can get C_m from the last equation, and then all coefficients are determined.

In order for a computer to solve the system of equations efficiently we write it on matrix form. We first note that we can ignore all D_i, except for D_0, due to the fact that we assume knowledge of the values σ_i. Therefore we will form block matrices where the different rows represent the equations above on the corresponding places. The columns will in general represent the coefficients of A_i, B_i and C_i.

We will write two matrices on a special form though. Since $p_0(s) = C_0 s + D_0$ we will skip the coefficients A_0 and B_0 entirely but have D_0 represented. For the same reason we will skip A_m and B_m. We define the block matrices,

$$a_1(p_0) := \begin{pmatrix} s_1 & 1 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$a_i(p_{i-1}) := \begin{pmatrix} (s_i - s_{i-1})^3 & (s_i - s_{i-1})^2 & (s_i - s_{i-1}) \\ 3(s_i - s_{i-1})^2 & 2(s_i - s_{i-1}) & 1 \\ 6(s_i - s_{i-1}) & 2 & 0 \end{pmatrix}$$

$$a_i(p_i) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -2 & 0 \end{pmatrix}$$

$$a_m(p_m) := \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}.$$

We insert them into a big matrix, denoted by V. The zeros in each column represent block matrices filled with zeros of the same dimension as the block matrices in that column that we have stated above. We define

$$V := \begin{pmatrix} a_1(p_0) & a_1(p_1) & 0 & . & . & . & . & . & . & 0 \\ 0 & a_2(p_1) & a_2(p_2) & 0 & . & . & . & . & . & 0 \\ . & & & & & & & & & . \\ . & & & & & & & & & . \\ 0 & . & . & 0 & a_i(p_{i-1}) & a_i(p_i) & 0 & . & . & . \\ . & & & & & & & & & . \\ . & & & & & & & & & . \\ 0 & . & . & . & . & . & . & 0 & a_m(p_{m-1}) & a_m(p_m) \end{pmatrix}.$$

In case there is any confusion, the zeros below $a_1(p_0)$ have the dimension

3x2, and the zeros above $a_m(p_m)$ have the dimension 3x1 while the rest of the zeros have the dimension 3x3. We thus have that V is a 3m-by-3m matrix. In order to completely structure up the system of equations we write down the coefficients of the polynomials as a vector. The same thing is done with the right hand side of the equations. We define O as the matrix containing all coefficients and L as the matrix representing the right hand side. We have

$$O := \begin{pmatrix} C_0 \\ D_0 \\ A_1 \\ B_1 \\ C_1 \\ \vdots \\ A_{m-1} \\ B_{m-1} \\ C_{m-1} \\ C_m \end{pmatrix} \qquad L := \begin{pmatrix} \sigma_1 \\ 0 \\ 0 \\ \sigma_2 - \sigma_1 \\ 0 \\ 0 \\ \vdots \\ \sigma_{m-1} - \sigma_{m-2} \\ 0 \\ 0 \\ \sigma_m - \sigma_{m-1} \\ 0 \\ 0 \end{pmatrix}$$

VO = L and we can solve for the coefficients in O by inverting V.

8 Reconstructing the volatility function

The minimizing problem and much of the notation follows the work in [3].

We now have at hand a method to produce the coefficients we seek when we know the values $(\sigma_1, \sigma_2, ..., \sigma_m)$. The problem is that we do not have access to these. We do however have access to market prices of options for different strike prices. We also have at hand the partial differential equation that, following our model of stock price movement, yields the option price, at least now that we have defined what our volatility function is. In section 5 we described how we are going to solve that partial differential equation numerically and ended up with the finite difference partial differential equation (30).

In section 6 we showed a method to fit our mathematical model to observed data, that is, we showed a method to solve a nonlinear least squares problem.

What is left to do is to define just what it is we are going to optimize, and what our scheme for doing this is.

8.1 The minimizing problem

Since we are going to do this with the help of a computer, and its computational capacity is limited, we will here let $s_{m+1} = 2s_{\text{init}}$, where s_{init} is the observed stock price corresponding to our observed option prices.

We will denote the interpolating cubic spline, described in section 7 as $c(s;\overline{\sigma})$, where $\overline{\sigma} := (\sigma_1, \cdots, \sigma_m)$, i.e. a vector containing the values that determines the coefficients of c. Thus,

$$c(s;\overline{\sigma}) := \begin{cases} p_0(s), & 0 \leq s < s_1 \\ p_1(s), & s_1 \leq s < s_2 \\ \vdots & \\ p_m(s), & s_m \leq s \leq s_{m+1}. \end{cases}$$

We use the same notation as when describing the Gauss-Newton method and denote the option value we get when solving the partial differential equation as $v_j(c(s;\overline{\sigma}))$, the meaning of the j is that this denotes the option value for a strike price K_j.

The parts that are unknown in this problem are the values in the vector $\overline{\sigma}$. In fact that is the only unknown part in our problem. It determines the values of the coefficients on the function we are trying to recreate from our data. Thus, the volatility values are the ones that will be optimized. We define an *inverse spline local volatility approximation problem*

$$\min_{\overline{\sigma}\in\mathbb{R}^m} F(\overline{\sigma}) = \frac{1}{2}\sum_{j=1}^{m}[v_j(\overline{v}_j - c(s;\overline{\sigma}))]^2, \tag{48}$$

where $\overline{v}_j$, just as before, denotes the observed market prices for options corresponding to a strike price K_j and m is the number of data we have at hand. The choice to have m number of data, just as the number of spline knots, is not a coincidence. More correctly would be to say that we choose to have as many spline knots as data we have at hand. When one normally uses splines it is to fit a model to observed points. Here we do not actually have data points that the *spline* has to go through, and we could therefore have chosen more (or fewer) spline knots than we had option prices. As noted in [3], having the same number of spline knots can lead to computational advantages so we choose that as our approach.

When looking for market prices on options it is typically difficult to find enough data on closed deals if one wants the same value on the underlying stock for every deal. The reason is partly that stocks are traded a lot, and thus the price changes often. Apart from that options are not traded that

much, and it can be hard to even find many closed deals with the same maturity, that is, during the same day.

There is however a way to solve this problem, at most times one can find a bid-ask pair on options for different strikes. A bid-ask pair $(\text{bid}_j, \text{ask}_j)$ consists of a *bid price*, denoted bid_j, which is the highest price someone is interested in paying for the option with strike price j at this moment. The *ask price*, denoted ask_j, is correspondingly the lowest price someone is willing to sell the option with strike price j for at this moment. We define the option price as

$$\overline{v}_j := \frac{\text{bid}_j + 3\text{ask}_j}{4}.$$

We weight it towards the ask prices based on a fairly small investigation on actual prices on options compared to the original bid-ask spread. It turned out to be in average this weight and therefore we use it here. The method used to get this weight is presented in appendix B.

If one has access to a more accurate approximation of the actual option price from bid-ask pairs then one could use that instead. It is also possible to instead use the last closed deal as the option price. However, we have chosen here to use one price on the stock for all strike prices, as well as one time to maturity. These restrictions forced us to take the option prices at one time for different strike prices, at least if we wanted more than a single option price to work with. The values of the options was taken after OMX closed for the day, making it possible to take the bid-ask pairs from the same point in time.

8.2 Setting up the numerical scheme

Note that this section is mainly a description of one way to set up a scheme using the methods we have described in the paper. We have not compared it to any other scheme, nor have we computed efficiency. It is just a suggestion as to how one could use the methods described in this paper to, hopefully, reconstruct a volatility function.

In order to start we will have to guess values on the vector $\overline{\sigma}$. We use the implied volatility to do this, that is, the volatility given by solving (6) for σ_i.

We are going to use the differential equation (30) to determine the current value of $v_j(c(s;\overline{\sigma}))$. We then need a way to update the values of $\overline{\sigma}$ in order to get a volatility function that is good enough for our minimizing problem. For this purpose we will use the Gauss-Newton method described in 6.

We assume that one has defined N and M, that is, defined at how many points one evaluates the partial differential equation. When one gathers option prices to fit the model to one gets access to the stock price s_{init} as well as strike prices K_j, $j = 1, \cdots, m$ and the time to maturity T. One also need the riskfree interest rate r. Since the vector $\overline{\sigma}$ is what will be updated we will denote the current vector with $\overline{\sigma}^{(k)}$. The scheme is the following:

1. Find coefficients to the spline $c(s; \overline{\sigma}^{(k)})$ by solving the system VO = L for your current vector $\overline{\sigma}^{(k)}$.

2. Calculate the theoretical option prices $v_j(c(s_{\text{init}}; \overline{\sigma}^{(k)}))$. Note that the option price is a function of the maturity time T, the stock price s_{init}, the riskfree rate of interest r, the strike price K_j and the volatility function c. We have chosen an approach where all but the strike price and volatility are the same. That is the reason that the option price function does not take all the other variables as variables, that is, we assume them as fixed.

3. Calculate $h_j(\overline{\sigma}^{(k)}) := \overline{v}_j - v_j(s_{\text{init}}; \overline{\sigma}^{(k)}))$ and define the vector

$$h(\overline{\sigma}^{(k)}) := \begin{bmatrix} h_1(\overline{\sigma}^{(k)}) \\ \vdots \\ h_m(\overline{\sigma}^{(k)}) \end{bmatrix}.$$

4. Calculate $F(\overline{\sigma}^{(k)}) = ||h(\overline{\sigma}^{(k)})||$. If $F(\overline{\sigma}^{(k)}) < \gamma$, where $\gamma > 0$ is defined to be the number deciding when we are "close enough", then we are done. Otherwise,

5. calculate

$$\nabla h(\overline{\sigma}^{(k)}) = \begin{bmatrix} \frac{\partial h_1}{\partial \overline{\sigma}_1^{(k)}} & \cdots & \frac{\partial h_1}{\partial \overline{\sigma}_m^{(k)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial \overline{\sigma}_1^{(k)}} & \cdots & \frac{\partial h_n}{\partial \overline{\sigma}_m^{(k)}} \end{bmatrix}.$$

   We have here tried to approximate the partial derivatives with

$$\frac{\partial h_i}{\partial \overline{\sigma}_j^{(k)}} \approx \frac{h_i(\overline{\sigma}^{(k)} + \epsilon e_j) - h_i(\overline{\sigma}^{(k)})}{\epsilon}.$$

   where $\epsilon > 0$ is chosen as small as allowed by Matlab, the program used to implement this. e_j represent a unit vector with the j:th element being equal to 1. It is likely that there are more accurate ways to calculate the derivatives that we have not looked into.

6. Solve the normal equation

$$(\nabla h(\overline{\sigma}^{(k)}))^T \nabla h(\overline{\sigma}^{(k)}) d = -(\nabla h(\overline{\sigma}^{(k)}))^T h(\overline{\sigma}^{(k)}) \tag{49}$$

   in order to,

7. update your vector $\overline{\sigma}^{(k)}$ and obtain a new one by $\overline{\sigma}^{(k+1)} = \overline{\sigma}^{(k)} + d^{(k)}$, where $d^{(k)}$ is the solution to the normal equation.

8. Repeat from 1.

References

[1] Baxter, M. and Rennie, A. *Financial Calculus: An Introduction to Derivative Pricing,* Cambridge University press, 1996.

[2] Black, F. and Scholes, M. The Pricing of Options and Corporate Liabilities, *The Journal of Political Economy, Vol. 81, No. 3 (May - Jun., 1973), pp. 637-654.*

[3] Coleman, T. F., Li, Y. and Verma, A. Reconstructing the Unknown Local Volatility Function, *Journal of Computational Finance 2(3), 1999, pp. 77-102.*

[4] Cornuejols, G., and Tütüncü, R. *Optimization Methods in Finance,* Carnegie Mellon University, Pittsburgh, USA (Summer 2005).

[5] Evans, L. C. *An Introduction to Stochastic Differential Equations Version 1.2,* UC Berkeley, *course notes.*

[6] Giles, M. B. and Carter, R. Convergence analysis of Crank-Nicolson and Rannacher time-marching, *Journal of Computational Finance, Vol. 9. No. 4 (Summer 2006), pp. 89-112.*

[7] Graham, R. L., Knuth, D. E. and Patashnik, O. *Concrete mathematics: a foundation for computer science,* Addison-Wesley Publishing Company, Reading, Massachusetts, second edition, 1994.

[8] Hull, J. C. *Options, Futures, And Other Derivatives,* global edition, Pearson Education Limited, eighth edition (2011).

[9] Merton, R. C. Theory of Rational Option Pricing, *The Bell Journal of Economics and Management Science, Vol. 4. No. 1 (spring 1973), pp. 131-183.*

[10] Morton, K. W. and Mayers, D. F. *Numerical Solutions of Partial Differential Equations: An Introduction,* second edition, Cambridge University Press, 2005.

[11] Recktenwald, G. W. *Finite-Difference Approximations to the Heat Equation,* Royal Institute of Technology, Stockholm, Sweden, (March, 2011), *course notes.*

[12] Rudin, W. *Principles of Mathematical Analysis,* McGraw-Hill Book company, New York, third edition, 1976.

[13] Sasane, A. *MA3260 Mathematical Economics,* Stockholm University, Stockholm, Sweden (July 2007), *course notes.*

[14] Sasane, A. and Svanberg, K. *Optimization (SF1811 /SF1831 / SF1841),* Royal Institute of Technology, Stockholm, Sweden, *course notes.*

[15] Weinstock, R. *Calculus of Variation with applications to physics and engineering,* Dover, 1974.

Appendices

A Assumptions in the normal Black-Scholes-Merton option pricing formula

The Black-Scholes-Merton formula uses a number of assumptions which are listed here, they all originate from page 309 in [8].

1. The stock price follows a geometric Brownian motion and the return μ aswell as the volatility σ are constant.
2. The short selling of securities with full use of proceeds is permitted.
3. There are no transactions costs or taxes. All securities are perfectly divisible.
4. There are no dividends during the life of the derivative.
5. There are no riskless arbitrage opportunities.
6. Security trading is continuous.
7. The risk-free rate of interest, r, is constant and the same for all maturities.

The difference in assumptions in this paper is that μ and σ does not have to be constant but functions of s and t, although we later on made the extra assumption that σ did not depend on t.

B Why $\overline{v}_j := \frac{bid_j+3ask_j}{4}$

This paper is mainly a mathematical one and data has been gathered with the only objective to have values to try and reconstruct a local volatility function. Due to this, the process of gathering data has not received much attention.

B.1 The method used to determine $\overline{v}_j = \frac{bid_j+3ask_j}{4}$

The method used to determine this weight was fairly simple and straightforward. We gathered all option prices from one day (from OMX), the same day the option prices were gathered, 2011-11-14, and chose every bid-ask pair that had a closed deal the last hour it was open. We then chose every bid-ask pair with a last closed deal that was higher than the bid price and lower than the strike price.

Let C_j denote the last closed deal on an option for a given strike price. For all these remaining pairs, the weight on each bid and ask was computed by solving the equation

$$w_1 \text{bid}_j + (1 - w_1)\text{ask}_j = C_j.$$

That is, we found out how close the last closed deal was to the current bid-ask pair. Once this was done, an arithmetic mean of all the weights on the bid prices was computed, and this resulted in the weights used here.