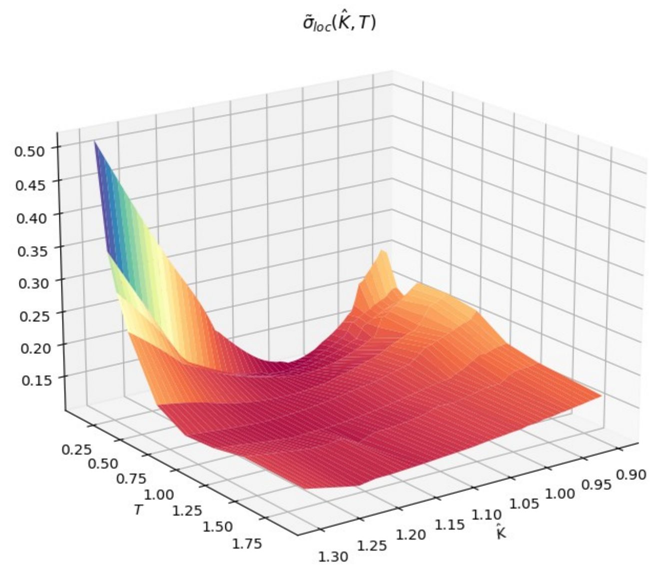
Degree Project in Applied and Computational Mathematics

Second cycle, 30 credits

Pragmatic Volatility Calibration

Numerical behaviour of the Andreassen-Huge Algorithm for calibrating the Dupire Local Volatility Model

OSCAR SPOLANDER



Abstract

In this work, grid selection in the Andreassen & Huge Algorithm (AHA) for calibrating Local Volatility was examined, providing the body of literature with a more detailed examination of the resilience of AHA to sparse grids. It was found that, if the goal is to calibrate a surface of European call option prices to market price data, AHA is capable of preserving model accuracy under grids much more sparse than those used in previous research.

Additionally, the plausibility of retrieving risk-neutral densities (RND) of the underlying asset using the AHA-derived call-surface was studied, which relies on a result from Breeden & Litzenberger [5]. These experiments, alongside the computations made when inverting call surfaces to Local Volatility surfaces, revealed that numerical differentiation of AHA-derived call surfaces is unstable. This was shown to be resolvable by selecting denser grids. Grid selection in the AHA algorithm thus depends highly on the goal of the practitioner, as well as their computational resource constraints.

Finally, a pricing scheme with quadrature integrals using the RND's and the AHA Local Volatility surface is proposed and evaluated against benchmark prices for the vanilla European call option on the Eurostoxx50 Index, and the digital European call option on the same underlying. The results suggest that this pricing scheme has some ability to accurately price instruments, while more research is neccessary to understand its capability to price a wider range of contracts.

Acknowledgements

I want to thank Boualem Djehiche, my supervisor at KTH, for his advice. His advice on narrowing the scope of the project and focusing on the essential at a key moment during the beginning phases of the research was especially valuable.

I also want to express my gratitude towards Jesper Tidblom of Algorithmica Research AB in Stockholm, as well as the entire organization at Algorithmica who welcomed my presence with open arms, offered me space in their office, and made this time truly joyful. Jesper spent a substantial amount of time and effort in mentoring my research project, which was priceless and something I will always be grateful for. Getting to learn from a true domain expert in quantitative finance has been a privilege.

Contents

Acronyms

AHA *Andreassen − Huge Algorithm*

BC *Boundary Condition*

DBE *Deutsche Börse Group Exchange*

ITM *In − the − money (option)*

IVC *Initial Value Condition*

LV *Local Volatility*

OTC *Over − the − counter (derivative)*

OTM *Out − of − the − money (option)*

PDE *Partial Differential Equation*

PNL *Profit − and loss*

RND *Risk − Neutral Density*

1 Introduction

Complete volatility models that fully capture the *whole* volatility surface given finite price data are important for pricing derivatives of generic strike/maturity combinations. In addition to computational budget constraints, practitioners also wish to adhere to theoretical conditions to ensure that their pricing models stay void of arbitrage on *all* points on the strike-maturity plane. This makes the calibration of volatility surfaces on a strike-maturity grid more intricate than merely applying the most computationally convenient interpolation that matches the price data; we need models that can quote fair prices on the *entire* dense plane.

In this work, we study stability and robustness in the Andreassen & Huge-algorithm (AHA) for calibrating the volatility surface. AHA is born out of the theory of so-called *Local Volatility* (LV) [3], which was an attempt to relax the Black-Scholes model as to accommodate empirical volatility smiles [10]. We also specify the theoretical constraints that ensure the absence of arbitrage for volatility surfaces, after which we address computational steps that practitioners can take to reliably create such surfaces for a wide range of pricing needs. This includes using the resulting volatility surface for pricing over-the-counter (OTC) derivatives for which data for calibration is unavailable.

As we shall see, the interpolation of market prices achieved by AHA via solving a PDE(Dupire's equation) numerically offers the interpretation of fitting a parameterized curve to data. In the AHA calibration, our curve has as many degrees of freedom as there are price data points. A statistician might consequently suspect problems with overfitting. Yet, such concerns lack relevance in local volatility calibration, since the ***objective*** of calibration is merely to deliver one of the multiple arbitrage-less surfaces of call prices. Whether it is consistent with some underlying ground truth call-surface is both impossible to infer and a non-goal for calibration.

In contrast, matters of greater relevance are the *conditions* under which AHA is able to deliver satisfactory call surfaces. By conditions, we here mean properties of the discretized grid for the strike-maturity plane, as well as the shape of input prices. Greater knowledge of such conditions could allow users of AHA to accomplish their goals with less computational resources, since the user-defined computational grid is adjustable by the practitioner (unlike the data-grid hosting the price data). Moreover, an understanding of limiting conditions beyond which AHA fails, can serve as a guideline for practitioners, alleviating them from rediscovering the wheel repeatedly.

This motivates us to direct the focus of our research to numerical vulnerabilities of AHA, as well as subsequent use cases of the model output. These include using the AHA volatility surface as inputs into other algorithms such as those used for pricing and hedging. We examine the stability of the finite difference solver for Dupire's equation at the boundaries on the strike-axis: a region where calibration algorithms are susceptible to collapse given flawed assumptions about asymptotic behavior in the tails [11] [24]. Also, we expose AHA to very sparse numerical grids and observe its tolerance to coarse interpolation. The current literature chooses to specify the grid rather heuristically, and given the importance of calibration speed for practitioners, we fill a gap in the literature by attempting to reduce computation time as a consequence of coarsening the discretization of Dupire's equation.

Finally, we attempt a pricing scheme using the implied transition density of the underlying: a quantity which can be obtained in the Dupire model from the so-called Breeden-Litzenberg formula, to be specified below in equation (10). Studying the reliability of this pricing algorithm is useful as it is viable for derivatives with any payoff, including exotic, illiquid instruments. Before we proceed with introducing background material, as well more topic-specific literature that has emerged from both industry and academia, we formally declare our multiple research questions below. Due to the ways in which the experiments that seek to answer these research questions depend on the background material, we present our experiment design under the section *method* further ahead in the report, which follows after the background section.

1.1 Research Questions

Rasmussen notes that AHA delivers unparalleled accuracy for calibrating market data since solving system (40) below, arising from Dupire's equation (22), is really a matter of interpolation [24]. Yet, it seems unrealistic to expect near-perfect calibration using grids of excessive sparsity. Take the limiting case $N_K^{numerical} = N_K^{data}$ [1] as a thaught-experiment. Then the finite difference solver of Dupire's equation has to adequately capture the second derivatives, the values of the prices, and the boundary conditions using a discretization only as dense as the data grid. It seems unreasonable to demand such capabilities from any interpolation algorithm. Meanwhile, the original paper [3] demonstrates successful matching to prices with fairly sparse grids of 100 strikes for 29 observed strikes per expiry. This is still far from the limiting case of $N_K^{numerical} = N_K^{data}$, which would have 29 computational nodes. This raises a research question of what

[1] Number of strikes in the numerical grid and data grid(data set) respectively.

an adequate numerical grid density really is in the AHA algorithm, especially since this appears to be rather unaddressed in the literature at present. We thus ask ourselves how sparse we can make the numerical grid while preserving acceptable calibration to data, where sparse grids offer savings in computational resources.

Moreover, though oscillating behaviour of AHA call surfaces near the **boundaries** has been mentioned in the literature [24], a closer examination of its cause and ways of mitigating these in a systematic manner is still missing. We attempt to close the gap in the literature by researching this matter more methodically. This issue is intrinsically tied to the first area of inquiry related to grid selection, since the width of the set of strikes [2] influences the size of the discretization matrix (number of grid points), which are in turn affected by the amount of buffering to support the far-field boundary conditions in (39). While the relationship between the Boundary Conditions(BVC) and buffering lies outside the scope of this study, we conjecture that the intuition practitioners use when employing buffering originates from how it enables solutions of Dupire's equation "room" to plateau towards the boundary conditions in (39) below , facilitating better reflections of true Dupire solutions.

Once model prices for European Call Options are retrieved using the AHA algorithm, we propose and test a pricing method for options with arbitrary payoffs using plain numerical quadrature with the implied density from the call surface, utilizing the fact that it coincides with the second derivative of the calls with respect to strike, as explained in equation (10) below. This integral is an attempt to estimate the risk-neutral expectation defining prices of contingent claims (financial derivatives), much like Monte-Carlo algorithms seek to compute expectation-integrals for options pricing [17]. As far as our literature inquiry went, we were not able to find accounts of using such pricing recipes, which motivates our inquiry into its efficacy. We formally specify our multiple research questions below

[2] We define this as the smallest strike on the grid subtracted from the largest one.

Research Question 1
Is there a threshold of grid sparsity in the discretization of Dupire's equation (22) beyond which AHA no longer achieves model call prices of acceptable accuracy? (1)

Research Question 2
What is the tightest possible buffering around the strike boundaries that preserves model accuracy across all strikes? (2)

Research Question 3
Is pricing of exotic derivatives with implied density and quadrature integration feasible and robust? (3)

1.2 Relevance

Seeking answers to the first question is important for practitioners with computational budget constraints, with time often being the most pressing one. A faster calibration procedure resolves what is otherwise a common bottleneck in the overall usage of local volatility models, and allows the user of such models to operate at a greater velocity.

Answers to the second question provide guidance to users of AHA on the matter of buffering in the endpoints, liberating algorithm users to rediscover the wheel each time.

The efficacy of generalized pricing with implied density would offer an important alternative to the pricing of general instruments including those with payoffs that still lack closed-form prices, like those of the Black-Scholes formula. Some claims are non-path-dependent, yet have payoffs that make expectation integrals intractable to evaluate analytically. These instruments sit somewhere in between path-dependent claims pricable with Monte-Carlo integration, and analytically tractable payoffs such as European call options in the Black-Scholes model, as far as implementation convenience is concerned.

2 Background

2.1 European call options

A conceptual understanding of volatility modeling requires basic familiarity with financial derivatives and arbitrage terminology, which we provide in these sections.

Definition 2.1 (European call option)**.** A **European call option** with price $C(K, T, s)$[3], strike $K > 0$, and maturity $T > 0$, written at time $t = 0$ on an underlying asset with a price-process S_t, with observed value $s := S_0$ at time $t = 0$, is a contract in which the holder enjoys the right, but not the obligation, to purchase the underlying asset at price K, at the future time T.

The option is thus worthless at time T if $S_T \leq K$, while it is worth $S_T - K > 0$ in case $S_T > K$. The option therefore has the *payoff function* $\phi(S_T) = \Big[S_T - K\Big]_+$[4] at T. Specifying *correct* prices $C(K, T, s)$ of European call options (and other financial derivatives) relies on arguments made about so called *arbitrage*; a situation where riskless profits are available on a financial market, implying the willingness of counterparties to offer free money. Our intuition tells us that rational traders reject such voluntary giveaways, forming the basis of the assumption that markets are *absent of arbitrage*, something we shall henceforth call ***price fairness*** or ***arbitrage freedom***, out of conciseness as to alleviate the need to repeatedly spell out *absence of arbitrage*.

Definition 2.2 (Arbitrage)**.** Formally, arbitrage is defined [4] as the presence, at some time t_0, of a trading position (collection of weights $\mathbf{h}_t \in \mathbb{R}^d$) on a financial market with $d \in \mathbb{N}$ number of assets collected in the vector stochastic process $\mathbf{S}_t \in \mathbb{R}^d$, characterized by

$$\mathbb{P}\left(\int_{t_0}^{\bar{t}} \mathbf{h}_t^T \mathbf{dS}_t \geq 0\right) > 0\ ,\ \ \mathbb{P}\left(\int_{t_0}^{\bar{t}} \mathbf{h}_t^T \mathbf{dS}_t < 0\right) = 0\ ,\ \ \mathbf{h}_{t_0}^T \mathbf{S}_{t_0} = 0\ , \tag{1}$$

for some $\bar{t} > t_0$, where the weights $\mathbf{h}_t$ are dynamically readjusted with no trading costs by a fictional investor in continuous time (i.e, able to make an

[3] We shall frequently allow ourselves to suppress notation into $C(K)$ or $C(T)$ out of conciseness, where context will hopefully and likely ensure unambiguity.

[4] The positive part, i.e $[x]_+ := max(x, 0)$.

infinite number of weight updates with no transaction costs). In the literature [4], the weights $\mathbf{h}_t$ are often called a "portfolio", and a **replicating portfolio** for the European call option with payoff $\phi(S_T) = \Big(S_T - K\Big)_+$ is one where

$$\int_{t_0}^{T} h_t dS_t = \phi(S_T) \qquad \mathbb{P} - \textit{ almost surely.} \tag{2}$$

2.2 Black-Scholes

While Local Volatility departs from the standard Black-Scholes model, their similarities put Black-Scholes among the prerequisites for understanding LV, which is why we present it here.

By the *Black-Scholes model* is meant [4] a trading environment offering investors and speculators an opportunity to allocate their wealth in either

1. A "risky" stock with a randomly time-varying value S_t
2. A "safe" bank account a with deterministic time-varying value B_t

whose evolution in time is governed by

$$\begin{aligned} dS_t &= rS_t dt + \sigma S_t dW_t \\ dB_t &= rB_t dt \ , \end{aligned} \tag{3}$$

where W_t is a Wiener process under the risk-neutral measure $\mathbb{Q}$; a time-continous stochastic process with properties described in texts such as [18] [12]. The interest rate r and the volatility σ in eq.(3) are both real constants under the model (3). Adopting interest rate models into the Black-Scholes framework adds randomness to r over time, while the elementary Black-Scholes model has constant interest rates [4]. We shall keep our interest rates constant in our study; a simplification we acknowledge comes with some inevitable model damage. The assumption made of constant(in time) volatility σ is the defining difference between the Local Volatility model used in this study, and the standard Black-Scholes model.

By "risk-neutral" measure $\mathbb{Q}$, we mean the probability measure[5] under which the normalized asset $Z_t := \frac{S_t}{B_t}$ is a martingale, i.e $\mathbb{E}^{\mathbb{Q}}\Big[Z_T \Big| \mathcal{F}_t\Big] = Z_t \quad , \quad T > t$. Using Itô's lemma [6], the stochastic differential equation (SDE) (3) can be solved

[5] For a definition, see for example, Durrett [12]

[6] See for example [22] for material on Itô's lemma

in closed form via the the change of variable $G_t := ln(S_t)$, so that

$$\begin{aligned}
&dG_t = \Big(r\frac{S_t}{S_t} - \frac{1}{2}\sigma^2\frac{S_t^2}{S_t^2}\Big)dt + \sigma\frac{S_t}{S_t}dW_t = \Big(r - \frac{1}{2}\sigma^2\Big)dt + \sigma dW_t \ . \\
&\text{Integrating the above yields} \\
&G_t = G_0 + \Big(r - \frac{1}{2}\sigma^2\Big)t + \sigma W_t \ , \ \text{ and it follows that} \\
&S_t = S_0 e^{\Big(r-\frac{1}{2}\sigma^2\Big)t+\sigma W_t} .
\end{aligned} \tag{4}$$

Note that, viewed at time $t = 0$, S_t is a random variable where *all* the randomness is caused by the Wiener process W_t in the exponent. Using the so-called Fundamental Theorems of Asset Pricing, equation (4) allows us to find a closed form expression of European call options. We present this procedure below.

2.3 Pricing

We now ask ourselves how financial institutions find fair prices of European call options in a Black-Scholes world. These options might be in demand by clients who wish to use such options as insurance against their vulnerability to future movements of the underlying asset price S_t. Out of conciseness, we do not go all the way back to derive the Black-Scholes equation and the subsequent risk-neutral pricing formula. Doing so relies on the absence of arbitrage inherent to the Black-Scholes model, as well as the application of Itô's lemma on the option price process $\Pi(t, S_t)$. Feynman-Kac's theorem is then used to conclude that the fair price $\Pi(t, S_t)$ of any financial derivative with payoff $\phi(S_T)$ is

$$\Pi(t, S_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\phi(S_T)\Big|\mathcal{F}_t\Big]. \tag{5}$$

For the European call option we find its price $C(K, T, S_t)$ by evaluating the right-hand-side of eq. (5) for $\phi(S_T) = \Big[S_T - K\Big]_+$, which one can derive to equate to

$$
\begin{aligned}
&C(K,T,S_t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\Big[\big(S_T - K\big)_+\Big|\mathcal{F}_t\Big] = \\
&= S_t\Phi\Big(d_1(t,S_t)\Big) - e^{-r(T-t)}K\Phi\Big(d_2(t,S_t)\Big)\ , \\
&d_1(t,u) = \frac{1}{\sigma\sqrt{(T-t)}}\Bigg[ln\Big(\frac{u}{K}\Big) + \big(r + \frac{1}{2}\sigma^2\big)(T-t)\Bigg]\ , \\
&d_2(t,u) = d_1(t,u) - \sigma\sqrt{(T-t)}\ .
\end{aligned} \tag{6}
$$

Where $\Phi(\cdot)$ is the standard gaussian c.d.f for $\mathcal{N}(0,1)$. Proofs of these results are readily available in the textbook by Björk [4].

2.4 Local volatility

While similar to the lognormal Black-Scholes model of the stock price dynamics in equations (3) and (4), the Local Volatility model relaxes the assumption made of constant volatility, in favor of a deterministic *volatility function* $\sigma(t, S_t)$ of time and the underlying stock itself [10]. The subsequent stock dynamics is then that of the Black-Scholes dynamics, but without constant volatility:

$$dS_t = rS_t d_t + \sigma(t, S_t) S_t dW_t. \tag{7}$$

We emphasize that all the randomness in $\sigma(t, S_t)$ arises from randomness in the stock, while the volatility itself is deterministic given known values of the input arguments t and S_t. Introducing separate sources of randomness into the volatility from that of the stock would lead us to a framework called *stochastic volatility*: another well-studied theory frequently used in industry [16]. Our work is limited to the study of local volatility, and for that reason a presentation of stochastic volatility is omitted in this report. A natural next step for the mathematician is now to specify $\sigma(t, S_t)$, in order to adequately price instruments under model (7). Local Volatility does not admit closed form-solutions of the terminal stock price, nor instrument prices such as those presented in (4)-(6) for the Black-Scholes model. However, it turns out that one can make significant progress towards specifying $\sigma(t, S_t)$ by differentiating call prices with respect to expiry and strike. This is what we shall do next, and the steps taken will hopefully become increasingly clear to the reader as we approach an expression for $\sigma(t, S_t)$. We first rewrite the price of call options as an integral starting from the strike K. Standing at time $t = 0$, prices of European call options are given by

$$e^{rT} C(K, T, s) = \mathbb{E}^{\mathbb{Q}}\Big[(S_T - K)_+ | \mathcal{F}_0\Big] = \int_K^\infty (x - K) p(x, T; s) dx. \tag{8}$$

In the above equation, $p(x, T; s)$ is what is known as the "transition density" of S_t from time 0 to T. This is the probability density of the random variable S_T, given information about the value s of S_0. To make it formal,

$$p(x, T; s) = \frac{\partial}{\partial x} \mathbb{Q}(S_T \leq x | S_0 = s). \tag{9}$$

As described by Gatheral [16], differentiating equation (8) twice with respect to K gives us

$$\frac{\partial}{\partial K} e^{rT} C(K,T,s) =$$

$$\int_K^\infty \frac{\partial}{\partial K}(x-K)p(x,T;s)dx = -\int_K^\infty p(x,T;s)dx \ ,$$

$$and\ for\ the\ second\ derivative :$$

$$\frac{\partial^2}{\partial K^2} e^{rT} C(K,T,s) = -\frac{\partial}{\partial K}\int_K^\infty p(x,T;s)dx = \tag{10}$$
$$-\frac{\partial}{\partial K}\left(\int_0^\infty p(x,T;s)dx - \int_0^K p(x,T;s)dx\right) =$$
$$\frac{\partial}{\partial K}\int_0^K p(x,T;s)dx = p(K,T;s) \ , \quad so$$

$$\frac{\partial^2}{\partial K^2} e^{rT} C(K,T,s) = p(K,T;s) \ ,$$

which is a result sometimes called the "Breeden-Litzenberg" formula [10], which will be useful soon. Permission to differentiate under integrals are justified in the appendix of Durrett [12]. We now differentiate equation (8) with respect to T, and the reason we do this is that $\sigma(t,S_t)$ then emerges, which was the variable for which we sought an expression in the first place. For any function $\phi(t,y) \in \mathcal{C}^{1\times 2}$, it is the case that

$$d\phi(t,S_t) =$$
$$\Big[\frac{\partial\phi(t,S_t)}{\partial t} + rS_t\frac{\partial}{\partial y}\phi(t,S_t) + \frac{1}{2}\sigma^2(t,S_t)S_t^2\frac{\partial^2}{\partial y^2}\phi(t,S_t)\Big]dt + \sigma(t,S_t)S_t\frac{\partial}{\partial y}\phi(t,S_t)dW_t \ ,$$

or written in integral form,

$$\phi(T,S_T) - \phi(t_0,S_{t_0}) =$$
$$\int_0^T \Big[\frac{\partial\phi(t,S_t)}{\partial t} + rS_t\frac{\partial}{\partial y}\phi(t,S_t) + \frac{1}{2}\sigma^2(t,S_t)S_t^2\frac{\partial^2}{\partial y^2}\phi(t,S_t)\Big]dt$$
$$+\int_0^T \sigma(t,S_t)S_t\frac{\partial}{\partial y}\phi(t,S_t)dW_t \ , \tag{11}$$

by virtue of Itô's lemma [22]. Taking conditional expectations of both sides, the diffusion term $U_t := \sigma(t, S_t) S_t \frac{\partial}{\partial y}\phi(t, S_t)$ vanishes [4] under sufficient integrability, leaving us with

$$\mathbb{E}\Big[\phi(T, S_T)\Big|\mathcal{F}_0\Big] - \phi(0, S_0) =$$
$$\mathbb{E}\Bigg[\int_0^T \Big(\frac{\partial\phi(t, S_t)}{\partial t} + rS_t\frac{\partial}{\partial y}\phi(t, S_t) + \frac{1}{2}\sigma^2(t, S_t)S_t^2\frac{\partial^2}{\partial y^2}\phi(t, S_t)\Big)dt\Bigg|\mathcal{F}_0\Bigg] =$$
$$\int_0^\infty \int_0^T \Bigg\{\Bigg[\frac{\partial\phi(t, y)}{\partial t} + ry\frac{\partial}{\partial y}\phi(t, y) + \frac{1}{2}\sigma^2(t, y)y^2\frac{\partial^2}{\partial y^2}\phi(t, y)\Bigg]\Bigg\}\, p\Bigg(y, T; s\Bigg)dtdy. \tag{12}$$

Assuming compact support of ϕ on $[0, T] \times \mathbb{R}_+$, along with adequate regularity conditions, one can, as Björk shows [4],

1. reverse the order of integration
2. apply integration by parts consecutively ,

to arrive at

$$\frac{\partial}{\partial t}p(x, T; s) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\sigma^2(t, x)x^2 p(x, T; s) - \frac{\partial}{\partial x}rxp(x, T; s)\ , \tag{13}$$

which is known as the *Fokker-Planck* equation. Our attention now returns to the call price in eq.(8), and in an attempt to extract the Local Volatility $\sigma(t, S_t)$, we differentiate the call price in eq.(8) with respect to T[7]:

$$\frac{\partial}{\partial T}e^{r(T-t)}C(K, T, s) = \int_K^\infty (S_T - K)\frac{\partial}{\partial T}p(S_T, T; s)dS_T \underset{eq.(13)}{=}$$
$$\int_K^\infty (S_T - K)\Bigg\{\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\sigma^2(T, S_T)S_T^2 p(S_T, T; s) - \frac{\partial}{\partial S_T}rS_T p(S_T, T; s)\Bigg\}dS_T. \tag{14}$$

Integrating by parts, acknowledging that $p(S_T, T; S_0)$ and its derivatives vanish at infinity, yields

[7] We assign credit to Gatheral's book here [16] for this approach.

$$
\begin{aligned}
&\int_K^\infty (S_T - K)\left\{\frac{1}{2}\frac{\partial^2}{\partial S_T^2}\sigma^2(T,S_T)S_T^2 p(S_T,T;s) - \frac{\partial}{\partial S_T} r S_T p(S_T,T;s)\right\} dS_T = \\
&\underbrace{\left[(S_T - K)\left\{\frac{1}{2}\frac{\partial}{\partial S_T}\sigma^2(T,S_T)S_T^2 p(S_T,T;s) - r S_T p(S_T,T;s)\right\}\right]_K^\infty}_{\text{zero, as } K-K=0 \text{ and } p \text{ and its derivatives vanish at } \infty} - \\
&\int_K^\infty \left\{\frac{1}{2}\frac{\partial}{\partial S_T}\sigma^2(T,S_T)S_T^2 p(S_T,T;s) - r S_T p(S_T,T;s)\right\} dS_T = \\
&\underbrace{-\left[\frac{1}{2}\sigma^2(T,S_T)S_T^2 p(T,S_T;s)\right]_K^\infty}_{\underset{eq.(10)}{=} \frac{1}{2}\sigma^2(T,K)K^2 e^{rT}\frac{\partial^2 C(K,T)}{\partial K^2}} + \quad \int_K^\infty r S_T p(T,S_T;s) dS_T = \\
&\frac{1}{2}\sigma^2(T,K)K^2 e^{rT}\frac{\partial^2 C(K,T)}{\partial K^2} \ + \underbrace{\int_K^\infty r S_T p(T,S_T;s) dS_T}_{:=\hat{I}}.
\end{aligned}
\tag{15}
$$

The motivation for taking all the derivatives is now more clear: we are getting closer to finding a functional form of $\sigma(t, S_t)$. We are left however, inconveniently, with the remaining term $\hat{I}$. Recall that this term arises from the drift $rS_t dt$ in the underlying stock SDE from eq.(7). Here, we take a very consequential step in order to liberate ourselves from dealing with $\hat{I}$ by removing the drift of equation (7) in the first place. We will see that this will then also remove $\hat{I}$. That next step is to ***normalize*** all prices by forward prices, which we describe next.

2.5 Normalization

We now normalize (divide) prices of the underlying asset and the strikes, by the so-called *Forward price*

$$F(t,T,S_t) := S_t e^{r(T-t)} \ , \tag{16}$$

which is the fair price of a contract to buy the asset at the future time T for its current price S_t. As mentioned, the reason for doing this is to remove $\hat{I}$ above in eq. (15), to find the local volatility $\sigma(t, S_t)$. Going forward, it is important to keep in mind that, standing at $t = 0$, $F(0, t, S_0)$ is *deterministic*. We now normalize (divide) the underlying stock process S_t by the (deterministic) forward value process $F(0, t, S_0)$, into the new (stochastic) process $X_t := \frac{S_t}{F(0,t,S_0)}$. Applying Itô's formula anew, we have

$$\begin{aligned} dX_t &= \underbrace{\Big[-rS_t + rS_t \Big]}_{0} \frac{e^{rt}}{S_0} dt + \hat{\sigma}(t, X_t) \underbrace{S_t \frac{e^{rt}}{S_0}}_{X_t} dW_t = \hat{\sigma}(t, X_t) X_t dW_t \ , \\ \hat{\sigma}(t, X_T) &:= \sigma\Big(t, X_t \cdot F(0,t,S_0)\Big) = \sigma(t, S_t) \ . \end{aligned} \tag{17}$$

We see in (17) that X_t is indeed a martingale Itô-process without drift (but with preserved diffusion term $\hat{\sigma}(t, X_t)X_t dW_t$) [8]. If one where now to derive the Fokker-Planck equation (13) but with X_t instead of S_t as the underlying process in Itô's lemma, the drift clearly vanishes in equation (11). Consequently, the first derivative with respect to K disappears for all intermediate steps in equations (11)-(13), and the Fokker-Planck equation with X_t as the underlying process thus becomes

$$\frac{\partial}{\partial t} p(x, T; X_0) = \frac{1}{2} \frac{\partial^2}{\partial x^2} \hat{\sigma}^2(t, x) x^2 p(x, T; X_0). \tag{18}$$

Moreover, the martingale property of X_t removes the necessity to factor the numeraire into fair normalized option prices written on X_t (the asset itself is already a martingale under $\mathbb{Q}$), so the discounting factor disappears. Before putting this back into the T-derivative in equation (14), we must also normalize

[8] The reader might notice how $\sigma(t, S_t)$ became $\hat{\sigma}(t, X_t)$, with possible skepticism. Note that they share values: the volatility $\hat{\sigma}(t, X_t)$ for X_t corresponding to S_t is equal to $\sigma(t, S_t)$.

the market call prices so that they align with the normalized price process X_t. Introducing the "normalized" call price $\hat{C}(T, \hat{K}, X_t,)$ with normalized strikes $\hat{K} := \frac{K}{F(t,T,S_t)}$, we see that

$$\hat{C}(\hat{K}, T, X_t) := \mathbb{E}\left[\left(X_T - \hat{K}\right)_+ \Big| \mathcal{F}_t\right] = \mathbb{E}\left[\frac{(S_T - K)_+}{F(t,T,S_t)} \Big| \mathcal{F}_t\right] =$$

$$\underset{F(t,\,T,\,S_t)\mathcal{F}_t\text{ - measurable}}{=} \quad \frac{1}{F(t,T,S_t)} \mathbb{E}\left[(S_T - K)_+ \Big| \mathcal{F}_t\right] = \tag{19}$$

$$\frac{e^{-rT}}{S_t} \mathbb{E}\left[(S_T - K)_+ \Big| \mathcal{F}_t\right] = \frac{1}{S_t} C(K, T, S_t).$$

This means that we can transition back and forth from normalized and regular call prices by mere multiplication (division) of the factor $\frac{1}{S_t}$, observable at t. It is thus clear that the Breeden-Litzenberg formula in (10) also holds for the normalized call prices:

$$\frac{\partial^2}{\partial \hat{K}^2} \hat{C}(\hat{K}, T, X_t) = p(x, T; X_0) \;, \tag{20}$$

where the transition density $p(x, T; X_0)$ of X_t is the scaled density $p(x, T; s)$ of S_t. Since no randomness is added by normalizing by forward prices,

$$\mathbb{Q}\Big(S_T \in [a - \epsilon, a + \epsilon]\Big) = \mathbb{Q}\left(X_T \in \left[\frac{a - \epsilon}{F(0,T,S_0)}, \frac{a + \epsilon}{F(0,T,S_0)}\right]\right) \tag{21}$$

are the same events in $\mathcal{F}_T$ for any $\epsilon > 0$. Repeating the steps in eqs. (8)-(15) on the normalized call prices with driftless X_t, the last integral in eq.(15) vanishes, and we get

$$\frac{1}{2} \hat{\sigma}^2(T, \hat{K}) \hat{K}^2 \frac{\partial^2}{\partial \hat{K}^2} \hat{C}\Big(\hat{K}, T, X_t)\Big) = \frac{\partial}{\partial T} \hat{C}(\hat{K}, T, X_t).$$

Rearranging the above gives us (22)

$$\hat{\sigma}^2(T, \hat{K}) = 2 \;\cdot\; \frac{\frac{\partial}{\partial T} \hat{C}(\hat{K}, T, X_t)}{\hat{K}^2 \frac{\partial^2}{\partial \hat{K}^2} \hat{C}(\hat{K}, T, X_t)} \;,$$

which is the renowned Dupire's formula for Local Volatility. Equation (22) above allows us to retrieve estimates of Local Volatility after obtaining call prices from using the Andreassen & Huge algorithm, as was highlighted in Rasmussen's work [24]. It was proposed that this is done using numerical differentiation with central differences:

$$\tilde{\sigma}_{loc}(T,\hat{K}) :=$$

$$\sqrt{2 \cdot \frac{\frac{\hat{C}(\hat{K},T+dT,X_t)-\hat{C}(\hat{K},T-dT,X_t)}{(T+dT)-(T-dT)}}{\hat{K}^2 \cdot \frac{\hat{C}(\hat{K}+d\hat{K},T,X_t)-2\cdot\hat{C}(\hat{K},T,X_t)-\hat{C}(\hat{K}-d\hat{K},T,X_t)}{[d\hat{K}]^2}}}\,, \tag{23}$$

which is what we do in this work in order to retrieve the volatility surface after calibrating the call surface with AHA. The non-uniform timesteps depending on T motivates the decision to write out the denominator of the numerator in (23), instead of the commonplace $2dT$ often seen in numerical analysis.

2.6 Arbitrage Conditions

Given the importance of preserving arbitrage absence in the call surfaces induced by volatility interpolation algorithms, we now present conditions imposed by the absence of arbitrage on prices of European call options. These constraints appear as inequalities, which emerge as a consequence of arbitrage absence [10]. In interpolation algorithms other than AHA, these inequalities enter naturally as constraints in the accompanying optimization problem [11] [24] [7]. Meanwhile, AHA sets itself apart from those algorithms since any set of call prices entering the loss function are inherently fair in AHA by virtue of its parametrization, i.e the functional form of call estimates chosen to model the market prices [3]. This entirely relaxes the constraints in the optimization procedure of AHA, resulting in an unconstrained optimization problem on $\mathbb{R}^d$.

As for the arbitrage conditions, a result from Carr and Medan ensures that eliminating so-called *Butterfly-spread arbitrage* and *Calendar-spread arbitrage* guarantees the absence of any static arbitrage [15]: a kind of arbitrage enjoyed with a position that is ***untouched*** until option expiry [6] (i.e, not rebalanced during intermittent times between acquisition of the hedging portfolio, and the expiry of the instrument being hedged). Therefore, solutions to the interpolation problem that enjoy freedom from these two kinds of arbitrage, are the kinds of solutions of importance in the literature, and are commonly also considered absent of arbitrage in general, rather than merely absent of static arbitrage.

We follow the footsteps of other researchers and also consider call surfaces void of butterfly arbitrage and calendar arbitrage to be plainly "free of arbitrage".

2.6.1 Butterfly Arbitrage

It follows naturally that prices of European call options in an arbitrage-less economy are monotonically negative with respect to strike; else one could simply sell the option of strike $K + dK$ and buy that with strike K, both expiring at time T to enjoy a riskless exposure of profits free of charge in any scenario where $K < S_T$. We thus have our first constraint for price fairness, in the form of monotonically decreasing call prices in terms of strike:

$$\frac{\partial C(K)}{\partial K} \leq 0 \quad , \quad \forall\, K \in \mathbb{R}_+ \,. \tag{24}$$

This provides us with some intuition for the origin of our next condition which is the convexity(in strike) imposed by the absence of so-called butterfly spread arbitrage; a trading opportunity that arises when said convexity is violated. We proceed to examine the second order derivative from a "trading perspective", which first requires a definition of the butterfly trade:

Definition 2.3 (Butterfly Spread)**.** A Butterfly Spread is a trading position with total price Π_{btf} in three different European Call Options, all expiring at time T, of three strikes priced at

$$C(K - dK)\ ,\ C(K)\ ,\ C(K + dK)\ , \tag{25}$$

with weights

$$\Pi_{btf} = C(K - dK) - 2C(K) + C(K + dK). \tag{26}$$

Irrespective of the call prices above, the *payoff* of the butterfly position is non-negative with full probability, implying a non-negative cost of the butterfly position on a fair[9] financial market. While this can be verified with some mental gymnastics on the part of the reader, we graph the butterfly payoff below:

[9]We take ourselves the liberty of using "fair" and "absence of arbitrage" interchangeably for conciseness.

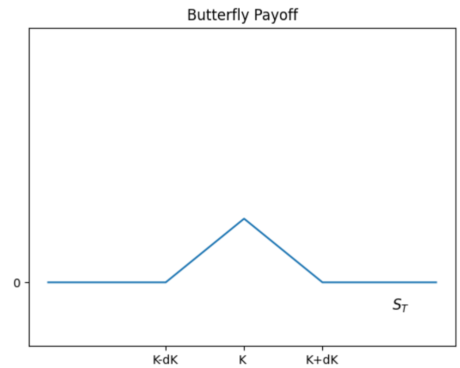


Figure 1: Butterfly payoff

Clearly the payoff is non-negative. Taking the limit $\lim\limits_{dK \downarrow 0}$ while acknowledging (26) then gives us the convexity condition imposed by absence of butterfly spread arbitrage:

$$\frac{\partial^2 C(K)}{\partial K^2} \geq 0 \; . \tag{27}$$

2.6.2 Calendar Spread Arbitrage

The definition of the Calendar Spread trade is analogous to the butterfly spread, except that the trading position has call options of different expiries instead of different strikes.

Definition 2.4 (Calendar Spread)**.** A Calendar Spread is a trading position in two different European Call Options, both with strike K, of two expiries T and $T + dT$, priced at

$$\Pi_{cal} = C(K, T + dT) - C(K, T). \tag{28}$$

Making use of the conditional Jensen's inequality on the convex function $f(x) = x_+$, one finds that, theoretically, call prices are monotone increasing in expiry:

$$
\begin{aligned}
&C(K,T+\Delta T) = e^{-r(T+\Delta T)}\mathbb{E}\Big[(S_{t+\Delta T}-K)_+\Big] = \\
&e^{-r(T+\Delta T)}\mathbb{E}\Big[\mathbb{E}\Big[(S_{T+\Delta T}-K)_+|\mathcal{F}_T\Big]\Big] \underset{Conditional\ Jensen}{\geq} \\
&\mathbb{E}\Big[\Big\{\mathbb{E}\Big[e^{-r(T+\Delta T)}(S_{T+\Delta T}-K)|\mathcal{F}_T\Big]\Big\}_+\Big] \underset{e^{-rt}\cdot S_t\ is\ a\ \mathbb{Q}-martingale}{=} \\
&e^{-rT}\mathbb{E}\Big[(S_T-e^{-r\Delta T}K)_+\Big] \geq e^{-rT}\mathbb{E}\Big[(S_T-K)_+\Big] = C(K,T).
\end{aligned}
\tag{29}
$$

Where we omitted the conditioning on $\mathcal{F}_0$ in this particular case to enhance clarity. To gain further intuition as to why violating eq. (29) would induce arbitrage on the market, imagine now entering the Calendar Spread position in (28) above at time $t = 0 < T < T+\Delta T$ if $C(K,T+dT) \leq C(K,T)$. The holder then owes $(S_T-K)_+$ upon the arrival of T, at which point they can fund their liability by selling the option maturing at $T+dT$ on the market for

$$
\begin{aligned}
&C(K,T+\Delta T) = \\
&e^{-r\Delta T}\mathbb{E}\Big[(S_{T+\Delta T}-K)_+|\mathcal{F}_T\Big] \underset{Conditional\ Jensen}{\geq} \Big\{\mathbb{E}\Big[e^{-r\Delta T}(S_{T+\Delta T}-K)|\mathcal{F}_T\Big]\Big\}_+ \\
&\underset{e^{-rt}\cdot S_t\ is\ a\ \mathbb{Q}-martingale}{=} (S_T-e^{-r\Delta T}K)_+ \geq (S_T-K)_+\ ,
\end{aligned}
\tag{30}
$$

which is exactly the liability facing the investor at time T as a consequence of entering the Calendar Spread in the first place. In the event of $C(K,T+dT) \leq C(K,T)$, it would then be the case that the market offers a free lunch since equation (30) implies a non-negative net total PNL of the Calendar Spread at time T. On a fair market void of arbitrage, it must therefore be the case that

$$
C(K,T+dT) \geq C(K,T)\ , \tag{31}
$$

which is the condition imposed by the absence of Calendar Spread arbitrage on the market for European call options. Carr [6] shows that eq. (31) and eq.

(27) are both ***neccessary and sufficient*** for the absence of static arbitrage. That is, once these inequalities hold, we can rest assured that our call prices are fair.

2.7 Calibration

The project of "calibrating Local Volatility" is to make an estimate $\hat{\sigma}^2_{model}(\hat{K}, T)$ of a supposedly "true" underlying local volatility function $\hat{\sigma}^2(K, T)$ given observations made on market data. AHA is not by any means the only method of doing this. Analogous to most endeavors of estimating reality via models, the practitioner makes a prior guess about the parametric form of the phenomena they try to capture, in our case $\hat{\sigma}^2(\hat{K}, T)$. We denote that guess $\hat{\sigma}^2_{model}(\hat{K}, T|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter vector in $\mathbb{R}^d$ for some parameter dimension $d \in \mathbb{N}$. For unit variable linear regression for example, we have $\hat{\sigma}^2_{model}(x|\boldsymbol{\theta}) = \theta_1 x + \theta_2$, $\boldsymbol{\theta} = \begin{bmatrix}\theta_1 \\ \theta_2\end{bmatrix} \in \mathbb{R}^2$ so $d = 2$.

Using an optimization algorithm, estimation then amounts to solving

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta} \in param\ space}{argmin} \ \mathcal{L}(\boldsymbol{\theta}) \ , \tag{32}$$

for some loss function $\mathcal{L}(\boldsymbol{\theta})$. In volatility calibration, practitioners choose variants of the weighted least-squares loss function [10]

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{N_E} \sum_{i=1}^{N_i} \frac{1}{w_{i,j}} (\hat{\sigma}_{model}(\hat{K}_{i,j}, T_i|\boldsymbol{\theta}) - \bar{\sigma}_{i,j})^2 \ , \tag{33}$$

for Black-Scholes implied volatilities $\bar{\sigma}_{i,j}$ from market data where the weights w_j assign different importance to errors depending on *normalized strike* $\hat{K}_{i,j}$. As mentioned earlier, the A&H algorithm distinguishes itself from other calibration methods in that the arbitrage conditions (27) and (31) are inherently satisfied due to its choice of $\hat{\sigma}^2_{model}(\hat{K}_{i,j}, T_i|\boldsymbol{\theta})$, enabling an unconstrained parameter space: *param space* $= \mathbb{R}^d$ since the arbitrage conditions are "already taken care

[10] In the sum below, N_E is the number of expiries in the data set, and N_i is the number of data points(prices) within expiry i, containing N_i strikes.

of". Many other algorithms impose the arbitrage conditions on the parameter space of (32), and carry out the optimization only in the part of $\mathbb{R}^d$ which satisfy the arbitrage constraints. We now proceed by outlining the shape of $\hat{\sigma}^2_{model}(\hat{K}, T|\boldsymbol{\theta})$ and $\mathcal{L}(\boldsymbol{\theta})$ in the particular case of the A&H-algorithm.

2.7.1 A&H Algorithm

For a dataset with N_E number of expiries $\{T_i\}_{i=1}^{N_E}$, each with N_i accompanying (normalized) call prices $\left\{\hat{c}(\hat{K}_{i,j}, T_i, X_0)\right\}_{j=1}^{N_i}$, the A&H-algorithm one expiry at a time places its full attention on ***each*** expiry T_i to solve the optimization problem in (33) ***individually*** for that expiry, before proceeding to the next expiry in the data set. For conciseness, we henceforth write $\hat{c}_{i,j} := \hat{c}(\hat{K}_{i,j}, T_i, X_0)$. In contrast to most other [11] [24] calibration schemes, A&H attempts to match model *call prices* to the market data:

$$\underset{\boldsymbol{\theta}\in param\ space}{argmin} \sum_{i=1}^{N_i} \frac{1}{w_{i,j}} (\hat{c}_{model}(\hat{K}_{i,j}, T_i|\boldsymbol{\theta}) - \bar{c}_{i,j})^2 = \tag{34}$$

$$\underset{\boldsymbol{\theta}\in param\ space}{argmin} \left|\left|\frac{\hat{\mathbf{c}}_{model}(\hat{\mathbf{K}}_i, T_i|\boldsymbol{\theta}) - \hat{\mathbf{c}}^i_{data}}{\mathbf{w}_i}\right|\right|^2 ,$$

whereas, as mentioned above, most other estimation algorithms calibrate in *volatility-space*:

$$\underset{\boldsymbol{\theta}\in param\ space}{argmin} \sum_{i=1}^{N_j} \frac{1}{w_{i,j}} (\hat{\sigma}_{model}(\hat{K}_{i,j}, T_i|\boldsymbol{\theta}) - \bar{\sigma}_{i,j})^2. \tag{35}$$

The intermediate model output from AHA is thus a set of call prices, after which one can retrieve local volatility from eq. (22) using numerical differentiation of the numerator and denominator as in (23), which we shall do and evaluate any subsequent errors emerging from numerical differentiation. To obtain model call prices $\hat{c}_{model}$, AHA discretizes Dupire's equation(22), resulting in a sparse tridiagonal system. To not conflate the data-grid described above with indices i, j with the numerical grid below, we now have the numerical indices k, q for the much denser (computational) strike-expiry grid. We address specifying the computational grid later.

A peculiar property of the grid is that previously published studies propose denser time steps in the numerical grid compared to the data grid [24], while elaborate experimentation led us to equate the data-grid with the numerical grid for the expiry axis. This is consistent with the original article [3], and means that the numerical grid underlying the model-retrieved surface is only denser on the strike axis, with more strikes on the numerical grid than the data grid, but with no more expiries. This is a finding that both limits what the model user can do with the model volatility surface (fewer strike/maturity combinations available for pricing exotics using the model surface), and raises questions for future researchers with respect to reproducibility. This result means that the q-indices and j-indices are the same in our implementation of AHA. We acknowledge that this digression arguably belongs in the section on results and/or conclusions. Meanwhile, we posit that this result has such significant effects on the design of the numerical grid that we decided that putting it here would enhance reader understanding and coherence. Ending this digression, we now proceed to specify the numerical solver of Dupire's equation. Combining central differences for the left hand side of eq.(22) with backward differences for the right hand side of the same equation, we get:

$$\frac{1}{2}\hat{\sigma}^2(\hat{K}_{k,q}, T_q)\hat{K}_{k,q}^2 \frac{\hat{c}_{k+1,q} - 2\hat{c}_{k,q} + \hat{c}_{k-1,q}}{\Delta \hat{K}^2} = \frac{\hat{c}_{k,q-1} - \hat{c}_{k,q}}{\Delta T}.$$

After rearranging and factorizing, we then get

$$\frac{1}{2}\hat{\sigma}^2(\hat{K}_{k,q}, T_q)\hat{K}_{k,q}^2 \frac{\Delta T}{\Delta \hat{K}^2}\Big[- \hat{c}_{k+1,q} + 2\hat{c}_{k,q} - \hat{c}_{k-1,q}\Big] + \Delta T \cdot \hat{c}_{k,q} = \hat{c}_{k,q-1} \ , \tag{36}$$

which represents a forward step in time (q-index). In order to start stepping forward from somewhere in the first place, we need an ***initial value condition***(IVC). Moreover, it turns out that particular behaviour deep in the money (ITM) or far out of the money (OTM) imposes a boundary condition that is critical for numerical stability. We describe both conditions next.

2.7.2 Initial Value Condition (IVC)

The IVC for solving (36) comes from the price of the normalized call option with expiry at $T = 0$, that is, with expiry today. Trivially, today's price of that contract with normalized strike $\hat{K}$ is its observable payoff. Noting that

$$X_{today} = \frac{S_{today}}{F_{today}} = \frac{S_{today}}{S_{today}} = 1\ , \tag{37}$$

we therefore get the IVC

$$\begin{aligned} \hat{c}_{k,0} &= \Big[1 - \hat{K}_k\Big]_+\ , \quad \text{or in vector form,} \\ \hat{\mathbf{c}}_{\cdot,0} &= \Big[\mathbf{1} - \hat{\mathbf{K}}\Big]_+ . \end{aligned} \tag{38}$$

2.7.3 Boundary Value Condition (BVC)

The A&H algorithm imposes *far-field* boundary conditions for the second-order derivatives in normalized strike $\hat{K}$: so called because they attempt capturing the limiting cases of $\hat{K} \uparrow \infty$ and $\hat{K} \downarrow 0$ [1]. A moment's reflection on the European call's payoff in these cases makes it clear that the payoff approaches 0 in the former case and $S_T - K$ with **full probability** in the latter. The second derivatives disappear in both ends, leading us to apply the boundary conditions

$$\frac{\partial^2}{\partial K^2}\hat{C}(K_{max}) = \frac{\partial^2}{\partial K^2}\hat{C}(K_{min}) = 0. \tag{39}$$

Since we use significant strike buffers on both ends, the probability of contract exercise has converged well within one and zero at K_{min} and K_{max} respectively, which provides the epistemic justification for applying the boundary conditions in (39). We are now in a position to specify the tridiagonal system resulting from equations (36) and (39), which amounts to

$$
\underbrace{\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -a_2^q & 1+2a_2^q & -a_2^q & \ddots & \ddots & \vdots \\ 0 & -a_3^q & 1+2a_3^q & -a_3^q & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -a_{N_q-1}^q & 1+2a_{N_q-1}^q & -a_{N_q-1}^q \\ 0 & \cdots & 0 & 0 & 0 & 1 \end{bmatrix}}_{=:A} \cdot \begin{bmatrix} \hat{c}_{1,q+1} \\ \hat{c}_{2,q+1} \\ \vdots \\ \hat{c}_{N_q-1,q+1} \\ \hat{c}_{N_q,q+1} \end{bmatrix} = \begin{bmatrix} \hat{c}_{1,q} \\ \hat{c}_{2,q} \\ \vdots \\ \hat{c}_{N_q-1,q} \\ \hat{c}_{N_q,q} \end{bmatrix}
$$

with

$$
a_k^q := 1 + \frac{1}{2}\hat{\sigma}^2(\hat{K}_{k,q}, T_q)\hat{K}_{k,q}^2 \frac{\Delta T}{\Delta \hat{K}^2}, \qquad \Delta T = T_{q+1} - T_q \; from\ data\ grid,
$$

$$
\Delta \hat{K} = constant,\ defined\ by\ the\ practitioner. \tag{40}
$$

The boundary condition from (39) enters the matrix A in the bottom and top rows. For the first time step when $q = 0$, the initial value condition (38) is used, after which the previous solution to (40) enables solving the subsequent steps.

2.7.4 Local Proxy Volatilities

In the above, the local volatility $\hat{\sigma}^2(\hat{K}_{k,q}, T_q)$ is written as a function, where we have not specified its form further. The critical insight made by Andreassen and Huge [3], was how a solution $\mathbf{\hat{c}_{q+1}}$ of (40) satisfies the previously mentioned arbitrage conditions if constant so-called ***volatility-proxies*** $v_{k,q}$ are used in place of $\hat{\sigma}^2(\hat{K}_{k,q}, T_q)$. This absence of arbitrage is a property also shared by linear interpolation between local volatility proxies, which is the setting chosen in our implementation. As highlighted by Rasmussen in her master's thesis [23], this very fact of preserving arbitrage freedom while simply solving a sparse linear system, underpins the popularity achieved by the A&H algorithm in industry, as practitioners now enjoy performance and price fairness all at once. For a given expiry indexed q, each step in an optimization scheme which we shall specify soon, computes the numerical gradients of

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}) &= \Big|\Big|\frac{\hat{\mathbf{c}}^{q+1}_{model} - \hat{\mathbf{c}}^{q+1}_{data}}{\mathbf{w}_q}\Big|\Big|^2 , \\
\hat{\mathbf{c}}^{q+1}_{model} &= A^{-1}\hat{\mathbf{c}}_{\mathbf{q+1}} ,
\end{aligned}
\tag{41}
$$

with respect to the local proxy volatilities $\theta_{k,q} := v_{k,q}$, and informs the optimization agent about how to best update $v_{k,q}$ as to improve(lessen) the loss function $\mathcal{L}(\boldsymbol{\theta})$. Our parameters $\boldsymbol{\theta}$ are thus volatility-proxies, and *one* iteration of the algorithm expiry-wise thus encompasses a back-and-forth process of solving the linear system in (40), finding the gradients of the loss $\mathcal{L}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, and evaluating $\mathcal{L}(\boldsymbol{\theta})$ after each update of the parameters(proxy vols) were made as to evaluate if the loss-criterion for terminating the algorithm has been found. Once it has, we pass the resulting parameters $\boldsymbol{\theta}_{optimal}$ into
$f(\hat{\mathbf{c}}_q;\boldsymbol{\theta}) := A^{-1}\hat{\mathbf{c}}_{\mathbf{q}}$ one last time to set our model output of the call-curve for that given expiry as

$$
\hat{\mathbf{c}}^{q+1}_{model} := f(\hat{\mathbf{c}}_q;\boldsymbol{\theta}_{optimal}). \tag{42}
$$

The invertability of A is ensured by its symmetry. Where the parameter-vector $\boldsymbol{\theta}$ enters in $f(\hat{\mathbf{c}}_q;\boldsymbol{\theta})$ is an understandable question a reader might ask, and we remind the reader that the volatility proxies(parameters) populate the A-matrix in (40). We now take this output $\hat{\mathbf{c}}^{q+1}_{model}$ as input into the right hand side of eq. (40) for the upcoming expiry and repeat the same(optimization) procedure N_E times. The result is a set of call prices on a dense grid on the strike-maturity plane, that can easily be interpolated linearly[11] to obtain a full call-surface: which to recall was the main goal for practitioners as stated in the introduction! Concerns of Smoothness(differentiability) of the call surface have been a frequently reported matter as practitioners seek to use the Breeden-Litzenberg formula (10) to study risk-neutral (probability) densities of the underlying asset. This motivates us to also discuss and conduct experiments related to the impact of various grids and interpolation schemes in the final step of retrieving transition densities; an artifact of relevance for various purposes in risk management and derivatives pricing [13]. The proof of price fairness of solutions $A^{-1}\hat{\mathbf{c}}_{\mathbf{q}}$ to eq (40) is lengthy, and we invite the reader to seek it out in the master's thesis of Rasmussen [23].

[11] Higher-order splines sadly introduces arbitrage here, as Lefloc'h notices [14].

For each normalized strike indexed k per fixed expiry of index q, the spacing of the local volatility proxies in the A&H-algorithm are set to

$$v_{k,q}(\hat{K}) = \begin{cases} \theta_1, & \text{if} \quad \hat{K} \leq \frac{\hat{K}_1^{market} + \hat{K}_2^{market}}{2}, \\ \\ \theta_k, & \text{if} \quad \frac{\hat{K}_{k-1}^{market} + \hat{K}_k^{market}}{2} < \hat{K} \leq \frac{\hat{K}_k^{market} + \hat{K}_{k+1}^{market}}{2}, \\ \\ \theta_{N_q}, & \text{if} \quad \hat{K} > \frac{\hat{K}_{N_q-1}^{market} + \hat{K}_{N_q}^{market}}{2}, \end{cases} \tag{43}$$

giving us N_q parameters (proxy volatilities) $\theta_{k,q}$ to calibrate for each expiry T_q, and a total of $N_q \times N_E$ parameters for the entire strike-maturity grid. As starting guesses for the proxy volatilities, the implied Black-Scholes volatilties are used, often retrivable from market data sets and if not, obtainable by inverting the Black-Scholes formula for European call options on market prices. A word on the optimization solver is warranted. While the optimization engine that solves (35) is a central part of the calibration algorithm, we have deliberately chosen to exclude matters of optimization from the study scope, in favor of the previously mentioned research questions (1)-(3), that pertain to numerical robustness. Rather, we rely on a well-trusted open-source implementation of a renowned optimization algorithm for non-linear optimization problems by the name of the *Levenberg-Marquardt Algorithm*. The open-source solver is provided by the sci-kit library in Python, which in turn relies on the *MINPACK*-library which was written in Fortran at the Argonne National Laboratory in 1980 and is widely deployed in production environments [21] [9]. Given the form of $\hat{c}_{model}(K, T|\boldsymbol{\theta})$, one can deduce that (35) defines a non-linear problem.

3 Method

3.1 Experiments

We seek answers to the stated research questions (1) - (3) by performing a number of experiments:

1. The first experiment is centered around the first research question (1) on grid selection for the finite difference solver of Dupire's equation. It encompasses running the entire algorithm multiple times with a consecutively sparser grid. We do this in order to identify a minimal grid density for which AHA can deliver on its promise of near-perfect interpolation (to

market prices). We then examine if there is a relationship between the numerical grid and the data grid. This is done by examining if data quantity and strike width[12] in a given maturity has an impact on grid selection, based on the hypothesis calibrating more data points with a given number of parameters (local volatility proxies) is more difficult than calibrating fewer points. In this first experiment, we use the same buffering scheme in the endpoints as that used by Rasmussen [24], which subtracts(adds) half the strike width(see footnote for definition) from the right(left) boundary respectively. This experiment yields insights into weather there is room to adjust the sparsity of the numerical grid based on that of the data grid, and serves the purpose of computational savings for the practitioner (we shrink the discretization matrix if we coarsen the numerical grid). Conducting these experiments thus provides the body of literature with novel knowledge of the practical use of AHA.

2. Our second experiment adresses buffering/padding at extreme strikes, meaning we ***deliberately insert*** extra gridpoints less than $\hat{K}_{min}$ and greater than $\hat{K}_{max}$ for stability reasons to do with the boundary conditions (39). This is of course intrinsically linked with grid sparsity: the central topic of experiment 1. What we try to find here is the presence (or absence) of a reliable scheme for buffering applicable to any incoming set of price data. The experiment plainly revolves around identifying constants ρ_1 , ρ_2 $\in \mathbb{R}_+$, but near one, such that the usage of numerical grids where

$$\left[\hat{K}_{min}^{numerical} \leq \rho_1 \cdot \hat{K}_{min}^{data}\right] \;\; AND \;\; \left[\hat{K}_{max}^{numerical} \geq \rho_2 \cdot \hat{K}_{max}^{data}\right] \tag{44}$$

implies stability in AHA,

that is, the algorithm displays stability only if the above buffering is used (for the specific data set). Stability here means that the optimization algorithm converges, which is indicative of sound intermittent solutions of $A\hat{\mathbf{c}}_{q+1} = \hat{\mathbf{c}}_q$ during the course of the optimization procedure. Again, researchers have thus far tended to pick ρ_1 and ρ_2 rather arbitrarily, subtracting(adding) whole big numbers to K_{min}^{data} and K_{max}^{data} [24]. This experiment seeks to find more specific insights into the matter, thereby alleviating users of AHA of the repeated guesswork involved in finding ρ_1 and ρ_2.

[12] With data quantity we mean the number N_j of data points for a given maturity indexed j, and strike width we define as $max(\{\hat{k}_{i,j}\}_{i=1}^{N_j}) - min(\{\hat{k}_{i,j}\}_{i=1}^{N_j})$ for the maturity j with N_j number of data points.

We do this by locating ρ_1 and ρ_2 via brute force for the various maturities, which can be seen as individual datasets, to see if any consistent patterns for ρ_1 and ρ_2 appear. Studying possible variations of ρ_1 and ρ_2 across maturities could offer insights into further optimization of grid selection because such variations could motivate AHA users to edit the grid *during* the algorithm for grid customization with respect to strike width (which is partly determined by the buffering). This is a seemingly unadressed question in the literature.

3. The third set of experiments focus on using the implied density, which the call surface makes available via the Breeden-Litzenberg formula (10). Retrieving the density has been deemed non-trivial by experts due to concerns about the smoothness of (computer)calibrated call-surfaces [20]. We put such claims to the test by differentiating the (model) call surface numerically with central differences, allowing us to observe the resulting (estimated) implied density:

$$\begin{aligned} &p_{S_T}^{estimate}(\hat{K},T;X_0) = \\ &\frac{\hat{C}_{model}(\hat{K}+\Delta\hat{K},T) - 2\hat{C}_{model}(\hat{K},T) + \hat{C}_{model}(\hat{K}-\Delta\hat{K},T)}{(\Delta\hat{K})^2} \approx \\ &\frac{\partial^2}{\partial\hat{K}^2}\hat{C}(\hat{K},T) = p_{X_T}^{true}(\hat{K},T;X_0). \end{aligned} \tag{45}$$

The discounting factor is absent in (45) due to price normalization, which removed the drift, as mentioned earlier. We will then attempt using the retrieved density to (numerically) compute pricing integrals

$$\begin{aligned} &Price_{financial\ derivative}^{Normalized} = \\ &\mathbb{E}\Big[\phi(X_T)|\mathcal{F}_0\Big] = \int_0^\infty \phi(x)p_{X_T}(x,T;X_0)dx\ , \\ &p_{X_T}(x,T;X_0) := \text{transition density of } X_T \text{ from 0 to } T, \end{aligned} \tag{46}$$

for derivatives with payoff functions $\phi(X_T)$ of the underlying (normalized) asset price X_T.

Our numerical integrator of choice is a plain quadrature rule

$$
\begin{aligned}
&Price^{estimate}_{financial\ derivative} = \\
&\sum_{i=0}^{N_{numerical}} \frac{1}{2}\Big[g(i\cdot\Delta x) + g((i+1)\cdot\Delta x)\Big]\cdot\Delta x \approx \\
&\int_0^{\infty}\phi(x)p_{X_T}(x,T;X_0)dx = Price^{true}_{financial\ derivative}\ , \\
&where\ g(x) := \phi(x)\cdot p^{estimate}_{X_T}(x,T;X_0).
\end{aligned}
\tag{47}
$$

Note that we use the grid with $N_{numerical}$ number of data points; which is the dense grid from the AHA calibration. Being aware that collecting the estimated implied density by numerical differentiation as well as computing price estimates from quadrature both introduce inevitable errors, we research if the approach in (45) - (47) at all has the capacity to solve pricing problems. Our first step in answering that question is to try to replicate the call prices originally derived from the market by computing $Price^{estimate}_{financial\ derivative}$ using the estimated implied density. If fruitful, we then proceed by attempting to price **digital options**[13] to further infer the capabilities that the above quadrature pricing method is endowed with.

We also use various methods for interpolating the dense model prices (linear, quadratic splines, cubic splines) and examine the subsequent smoothness they induce on $p^{estimate}_{S_T}(x;t,X_t)$, given the intuition that the second derivative bears relevance to the smoothness of the call surface, as emphasized by LeFloc'h [14] and revealed when differentiating the said surface twice.

Surprisingly, while much has been written about the utility of retrieving the implied density from call surfaces [13], the practice of directly using the density to price instruments numerically appears to be scarce, if not absent, in the current literature.

[13]options with payoffs of the kind $\phi(S_T) = S_T\cdot\mathbb{1}_{S_T>K}$.

3.2 Data

We leverage the internet to retrieve price data from the Deutsche Börse Group Exchange (DBE) [8]. DBE maintains what they call a "Data Shop" where researchers can access data. A free example dataset of Eurostoxx 50 (SX5E) with variable expiry-gaps, with expiries up to 2.88 years, was retrieved from the web. An important detail of the SX5E data is that it contained precomputed values of the greeks, meaning that we rely on external data for the market-derived vegas which enter as the weights in the loss function for eq.(33). Additionally, implied Black-Scholes volatilities had been precomputed in the data set, which appear in the optimization solver as starting-guesses for the volatility proxies (parameters). We remain confident in the reliability of these pre-computed values, given the reputational weight of the institution behind these datasets. We also doubled-checked their validity by inverting the implied volatilities back to call prices ourselves, and got identical quantities.

3.3 Error Reports

To measure estimation quality of call prices coming out of AHA, we use a metric slightly different from the loss function in (33) of the optimization problem. This is partly due to the fact that the weights w_i distorts the measure, albeit to a minuscule extent. Moreover, the final model prices are are not the exact same values as the prices of the last iteration of the Levenberg-marquardt optimizer, since the final prices are obtained by computing
$\hat{\mathbf{c}}^q_{model} = f(\hat{\mathbf{c}}_q; \boldsymbol{\theta}_{optimal}) = A^{-1}\hat{\mathbf{c}}_{\mathbf{q-1}}$ one last time to obtain the ultimate estimates $\hat{\mathbf{c}}^q_{model}$. Consequently, we define our *Mean Square Error* by comparing those values against the original model prices:

$$Mean\ Square\ Error = \frac{1}{N_q}||\hat{\mathbf{c}}^q_{model} - \hat{\mathbf{c}}^q_{market}||^2 \,, \tag{48}$$

where N_q is the number of data points in the data set for expiry q.

4 Results

4.1 Grid Sparsity

We found that accurate call prices can be obtained with grids that were significantly more sparse than those used in previous studies. With up to 38 strikes per expiry, we were able to use less than twice that amount of grid points when solving Dupire's PDE while matching price data to the degree seen in the upper left plot below, with a mean square error of at most $1.28e-07$, resulting from the expiry at 0.041 years, containing 38 strikes:

Figure 2: **Normalized european call Eurostoxx50 index option prices from the AHA algorithm with $N_{numerical} = 60$ and buffering** $\frac{1}{2} \cdot \textit{strike width}$

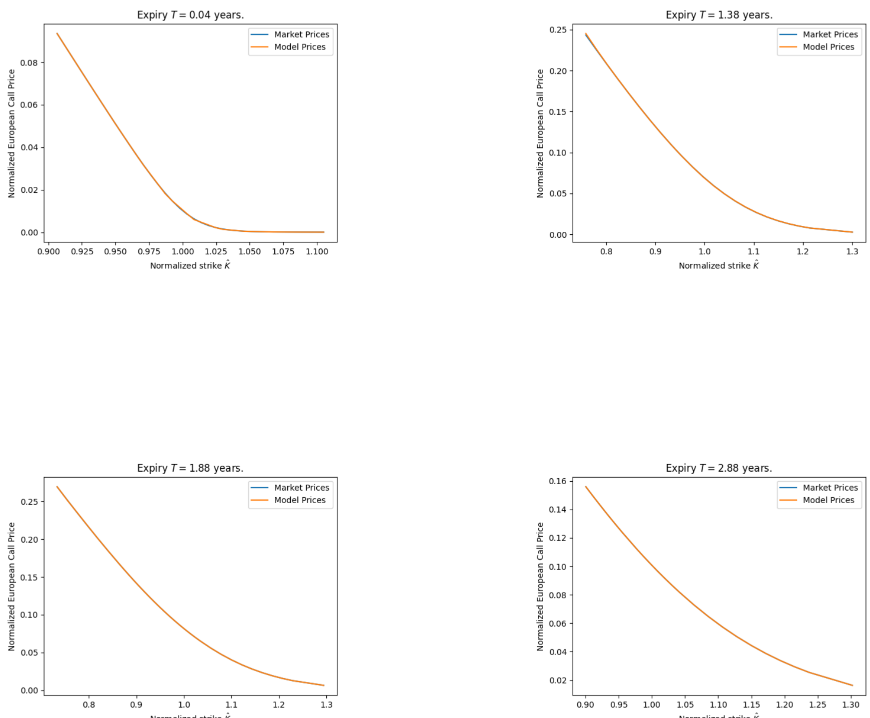


Table 1: **Errors for AHA model prices with with** $N_{numerical} = 60$ **and buffering** $\frac{1}{2} \cdot \textit{strike width}$

Expiry	Total Error	Mean Square Error	Max Error	Num Data Points
4.11e-02	5.07e-07	1.33e-08	1.28e-07	38
1.18e-01	2.64e-07	5.62e-09	3.07e-08	47
2.14e-01	1.96e-07	3.38e-09	1.96e-08	58
2.90e-01	2.28e-07	3.68e-09	3.17e-08	62
3.86e-01	1.32e-07	1.94e-09	1.03e-08	68
6.36e-01	4.76e-08	1.25e-09	5.15e-09	38
8.85e-01	3.31e-08	8.48e-10	4.38e-09	39
1.13e+00	3.07e-20	1.39e-21	1.56e-20	22
1.38e+00	4.44e-06	1.85e-07	3.60e-06	24
1.88e+00	1.07e-07	3.23e-09	4.00e-08	33
2.88e+00	3.92e-09	1.70e-10	8.54e-10	23

Further coarsening of the normalized strike axis to 40 uniform points over the same interval $\frac{K}{F_t(T)} = \hat{K} \in [0.3666, 2.604]$ caused minor damage to estimation quality, while even this grid selection gave us relatively modest errors. Notably, the visual effects of this sparsening of the grid one can argue is somewhat deceptive, where the actual errors reported in the table below remain small while the plot for the first expiry appears to suffer from visible misfits. This is likely due to the model matching the price data while placing no priority on any estimation in the intermittent points, which has been debated in the literature and which we will discuss in section 5.1 under discussion.

Figure 3: **AHA model normalized european call Eurostoxx50 index option prices with with** $N_{numerical} = 40$ **and buffering** $\frac{1}{2} \cdot \textit{strike width}$

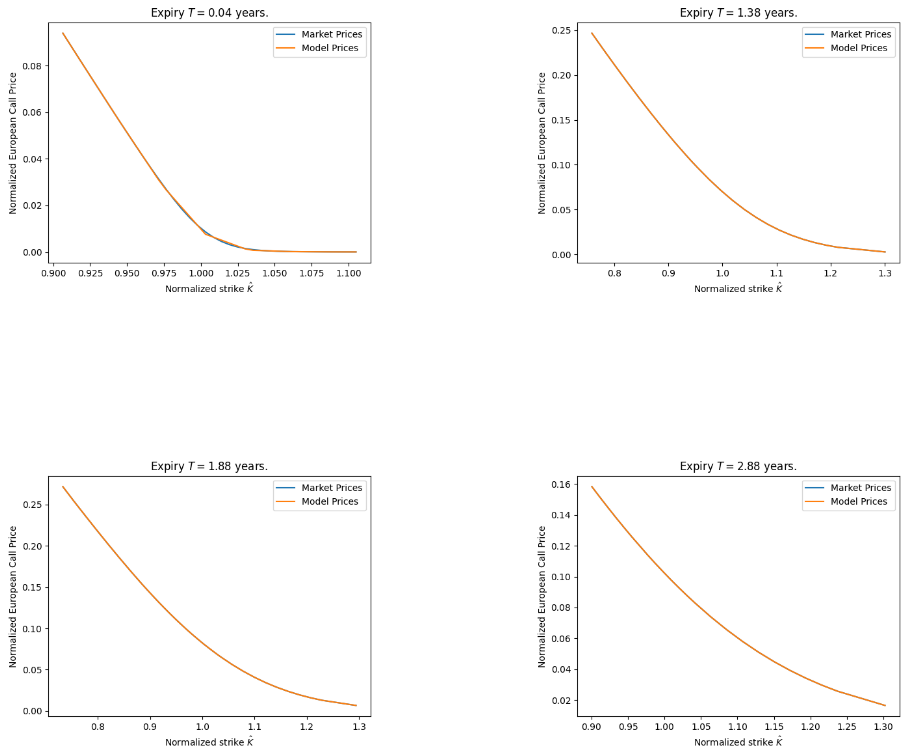


Table 2: **Errors for AHA model prices with with** $N_{numerical} = 40$ **and buffering** $\frac{1}{2} \cdot \textit{strike width}$

Expiry	Total Error	Mean Square Error	Max Error	Num Data Points
4.11e-02	3.11e-06	8.19e-08	1.02e-06	38
1.18e-01	1.47e-06	3.13e-08	2.34e-07	47
2.14e-01	1.04e-06	1.78e-08	1.33e-07	58
2.90e-01	8.51e-07	1.37e-08	1.05e-07	62
3.86e-01	7.03e-07	1.03e-08	7.13e-08	68
6.36e-01	2.75e-07	7.23e-09	3.23e-08	38
8.85e-01	2.30e-07	5.90e-09	1.88e-08	39
1.13e+00	8.42e-09	3.83e-10	3.22e-09	22
1.38e+00	4.58e-06	1.91e-07	3.86e-06	24
1.88e+00	2.51e-07	7.60e-09	7.41e-08	33
2.88e+00	2.70e-08	1.17e-09	3.28e-09	23

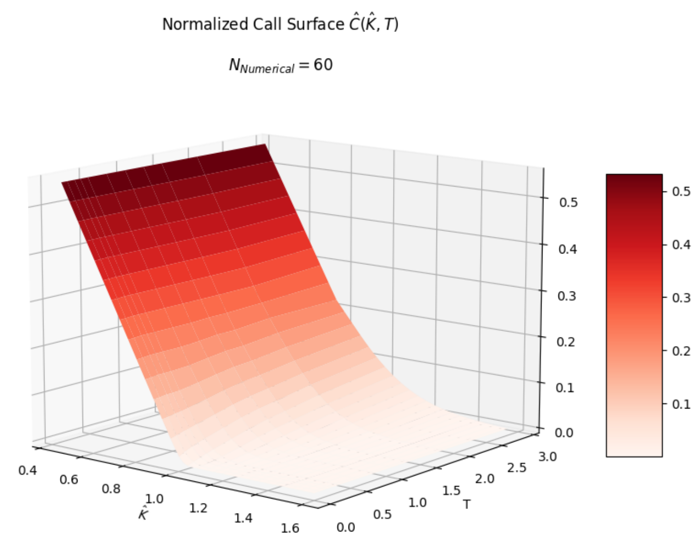


Figure 4: Sample Call surface

4.2 Boundary Conditions

Adjusting the buffering around the boundaries had a much more adverse effect on model quality than that of the aforementioned sparsening of the grid. Even with a much denser grid of $N_{numerical} = 10,000$, we saw significant errors at the boundaries when removing buffering entirely. This test was done to eliminate the possibility of grid spacing causing this error-prone behavior. Notably, these errors appear to scale as you move out expiry-wise:

Figure 5: **AHA model prices with no buffering at far strikes, using** $N_{numerical} = 60$

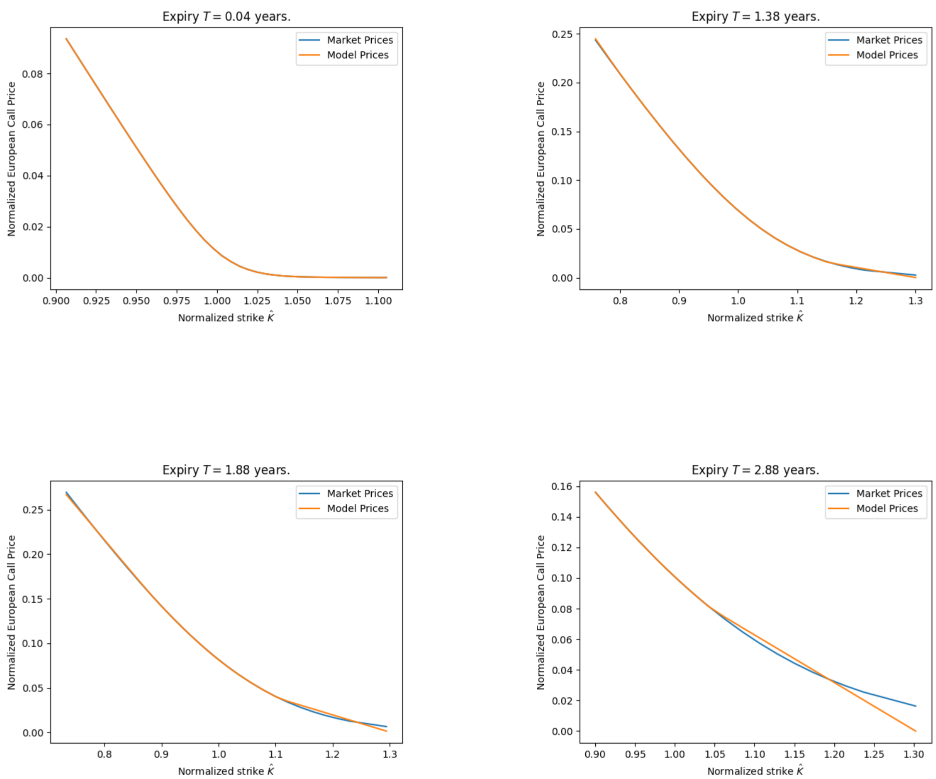


Table 3: **Call Price errors with no buffering at far strikes for** $N_{numerical} = 60$

Expiry	Total Error	Mean Square Error	Max Error	Num Data Points
4.11e-02	3.20e-08	8.42e-10	7.26e-09	38
1.18e-01	1.38e-08	2.94e-10	1.99e-09	47
2.14e-01	1.05e-08	1.81e-10	1.26e-09	58
2.90e-01	8.43e-09	1.36e-10	9.45e-10	62
3.86e-01	7.17e-09	1.05e-10	6.91e-10	68
6.36e-01	3.05e-10	8.03e-12	1.11e-10	38
8.85e-01	1.99e-20	5.11e-22	2.83e-21	39
1.13e+00	8.28e-07	3.76e-08	6.16e-07	22
1.38e+00	1.29e-05	5.39e-07	6.38e-06	24
1.88e+00	6.90e-05	2.09e-06	2.55e-05	33
2.88e+00	3.45e-04	1.50e-05	2.66e-04	23

Note how the calibration errors appear at the far strikes. Also, the overall residuals increase for longer expiries, and notably, in the longest expiry, the maximum error is responsible for nearly all of the total error. A simple explanation for this result is that the expiries each has different spans of strikes, so removing the buffering for the span of total observed strikes in the data still provided a buffer for the expiries with minimum and maximum well within the total min/max.

Yet, this does not entirely explain the cause for poor calibration near the boundaries, since the options with expiries $T = 1.384$, $T = 1.882$ and $T = 2.879$ had the same greatest normalized strike of 1.30, so there appears to be a link between this problem and the time horizon(expiry) as well, not just a lack of buffering in the right endpoint. After trial and error, we found that for the particular data set for the Eurostoxx50 index options, a buffering of $\rho_1 = \frac{1}{1.1}$ and $\rho_2 = 1.1$ where tolerable; vastly lower than the buffering reported in previous research, resulting in significant computational savings. Notably, using $\rho_1 = 0.95$ and $\rho_2 = 1.05$ gave satisfactory calibration for all but the last expiry, despite the fact that it had the same greatest strike as the previous expiries.

4.3 Risk-Neutral Densities

We saw above that Local Volatility estimation with AHA is very robust even under sparse computational grids. We carried out similar experiments for density estimation, where we observed the effect of grid selection on RND estimates. We retrieved risk-neutral densities (RND) under four different conditions:

A. Dense AHA grid using the dense numerical grid for central-difference differentiation.

B. Dense AHA grid using the coarser data grid for central-difference differentiation.

C. More sparse AHA grid using the numerical grid for central-difference differentiation.

D. More sparse AHA grid using the even coarser data grid for central-difference differentiation.

Where we define a dense numerical grid as that containing 2000 strikes, and our coarse numerical grid as that containing 80 strikes. In these experiments, we used the buffer constants $\rho_1 = \frac{1}{1.1}$ and $\rho_2 = 1.1$ as before. One can visually evaluate the resulting RND's below:

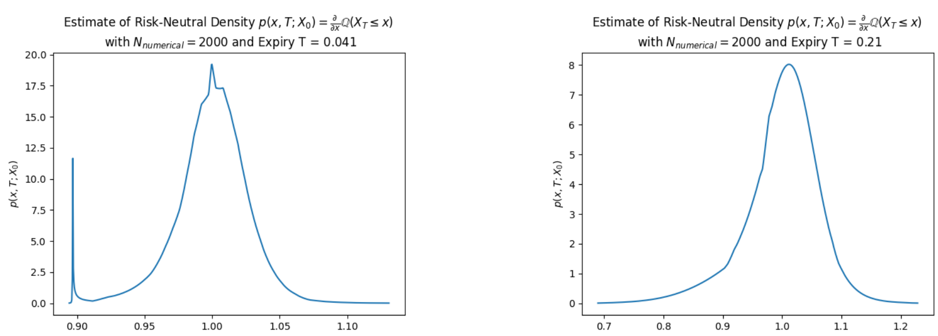


Figure 6: A. Implied Densities with dense numerical grid, using numerical grid for differentiation.

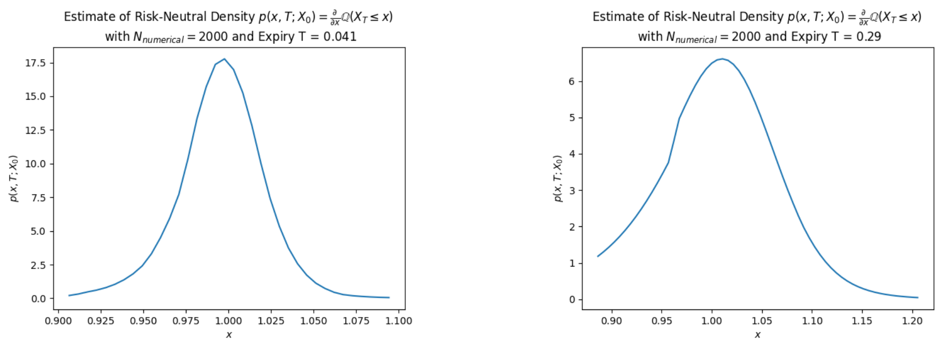


Figure 7: B. Implied Densities with dense numerical grid, using data grid for differentiation.

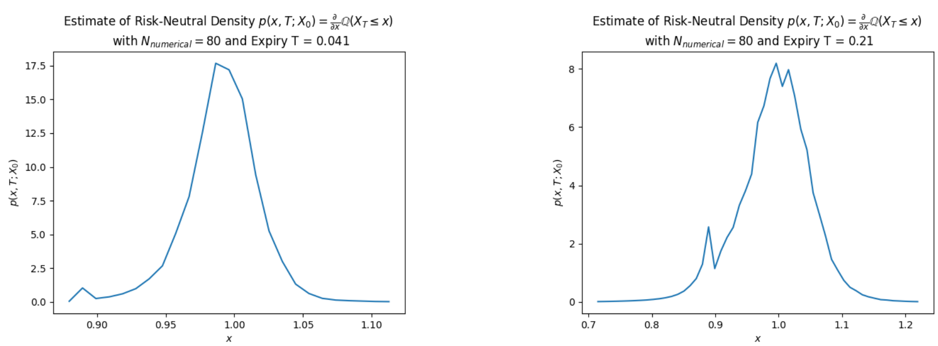


Figure 8: C. Implied Densities with coarse numerical grid, using numerical grid for differentiation.

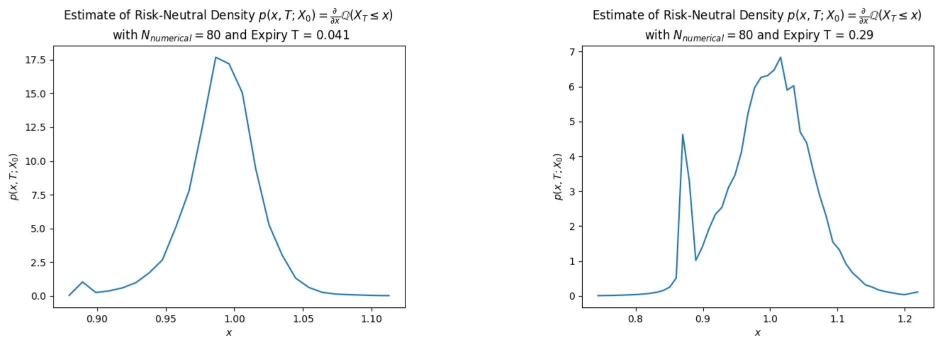
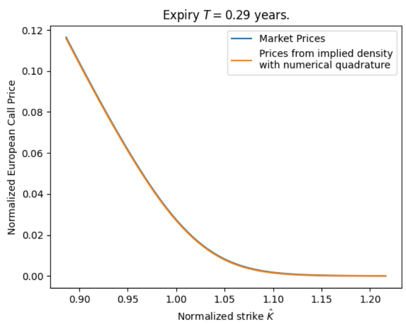
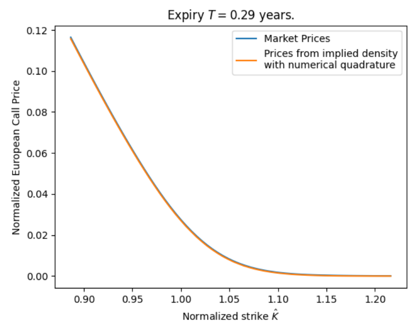


Figure 9: D. Implied Densities with coarse numerical grid, using data grid for differentiation.

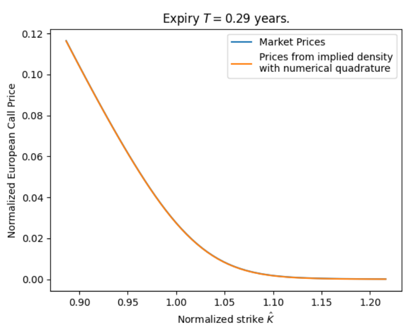
Using the numerical quadrature pricing scheme, these densities then reproduced call prices adequately for the expiries with satisfactory densities. We below plot prices for $T = 0.29$, for schemes A-D, after which we address problems appearing further out in time.



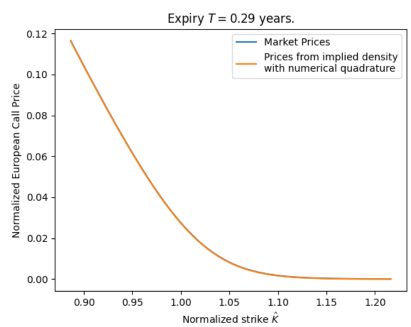
With density from A



With density from B



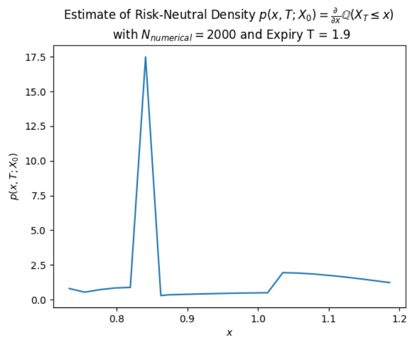
With density from C



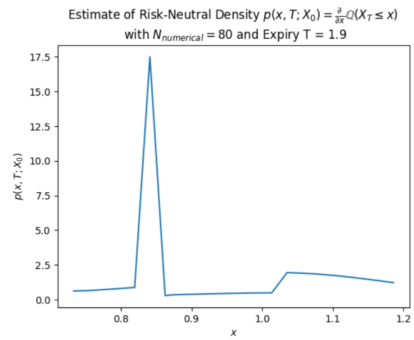
With density from D

Figure 10: Reproducing market call prices using risk-neutral densities with quadrature pricing.

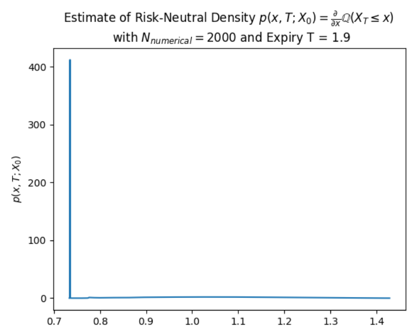
Our most notable result studying RND estimation is that all schemes collapse for longer maturities:



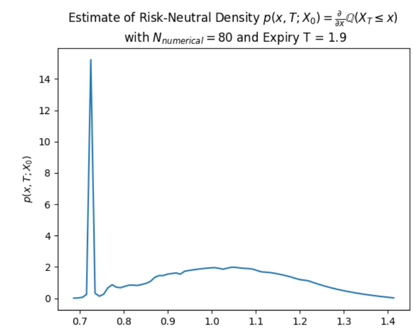
With density from A



With density from B



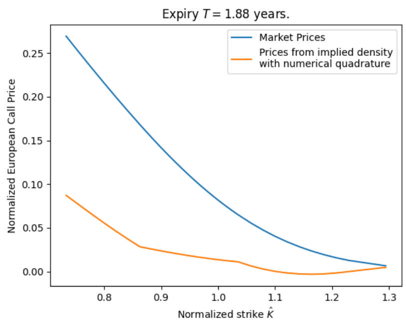
With density from C



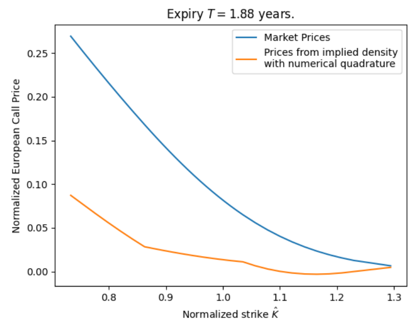
With density from D

Figure 11: Flawed risk-neutral densities

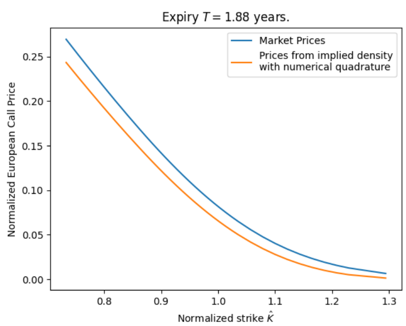
Resulting in inaccurate pricing using quadrature integrals:



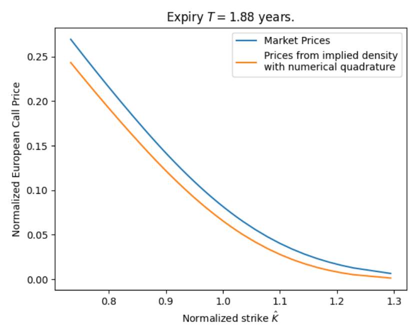
With density from A



With density from B



With density from C



With density from D

Figure 12: Pricing errors using flawed densities with quadrature pricing

The below table provides an overview of the results for *all* expiries, where one can notice the rapid decline of estimate qualities in this method for long maturiries, in line with above plots:

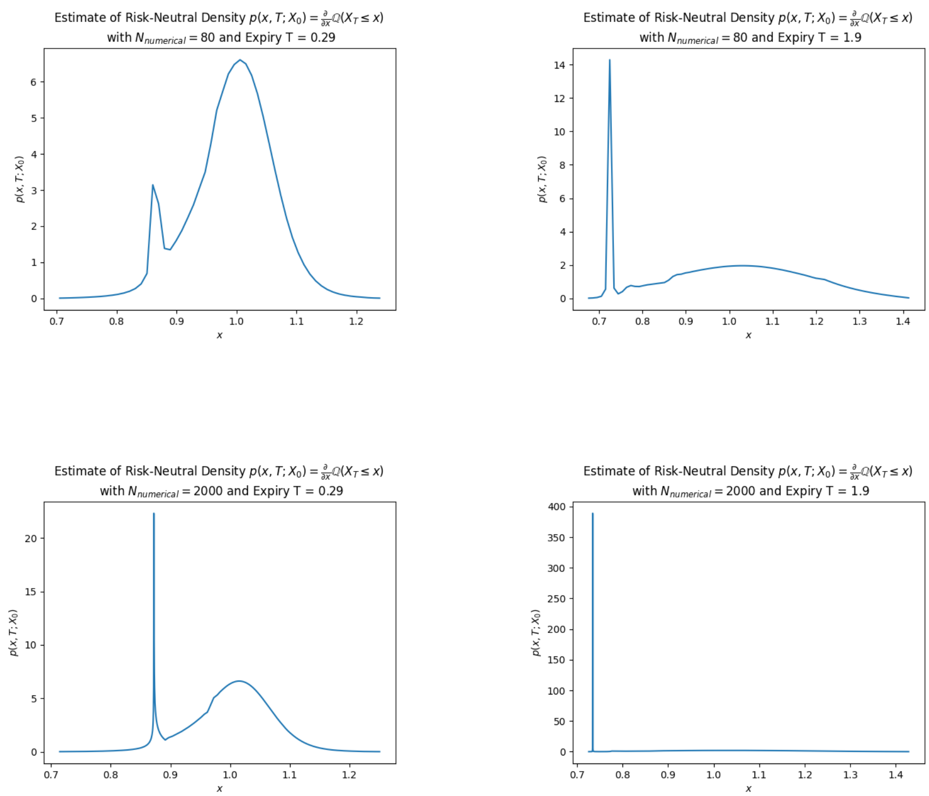
Table 4: **Errors for quadrature pricing with RND of Vanilla European call options on the Eurostoxx50 index**

Expiry	Total Error	Mean Square Error	Max Error	Num Data Points
4.11e-02	9.24e-09	2.43e-10	3.63e-10	38
1.18e-01	3.91e-08	8.32e-10	1.37e-09	47
2.14e-01	2.22e-07	3.83e-09	7.10e-09	58
2.90e-01	6.30e-07	1.02e-08	1.96e-08	62
3.86e-01	1.27e-06	1.87e-08	3.79e-08	68
6.36e-01	6.12e-06	1.61e-07	3.18e-07	38
8.85e-01	3.46e-05	8.88e-07	1.80e-06	39
1.13e+00	1.67e-04	7.59e-06	1.55e-05	22
1.38e+00	9.93e-04	4.14e-05	6.88e-05	24
1.88e+00	1.01e-02	3.07e-04	6.76e-04	33
2.88e+00	5.47e-02	2.38e-03	4.36e-03	23

While interpolating[14] with splines does not preserve arbitrage in the AHA call surface, one could think of a case where a practitioner is interested in solely studying implied densities of the underlying: a goal which is independent of preserving arbitrage. This affords an opportunity to use interpolation with cubic splines to induce thrice differentiability in an attempt to achieve a $\mathcal{C}^3$-call surface with smoother second order derivatives, and consequently, risk-neutral densities. Given the superior density estimates of using the numerical grid for central differences seen above, we only ran RND retrieval with cubic spline interpolation using that setting. Doing this resulted in the following results, picking one early/late arbitrarily chosen expiry each:

[14] The call prices, not the proxy-volatilities. Interpolation appears in both of these steps within AHA.

Figure 13: Implied Densities with cubic spline interpolation for the call surface



We clearly see that this approach does not resolve RND collapse for long maturities, while one can observe smoother forms for the densities that did more adequately model true densities.

4.4 Local Volatility Retrieval

Using the AHA model call surface, equation (22) was inverted to obtain a corresponding local volatility surface implied from the European call options. Central differences were used to approximate the derivatives, as proposed in the Phd-thesis from Rasmussen [24] and specified in eq.(23). One can visualize local volatility in the figures below:

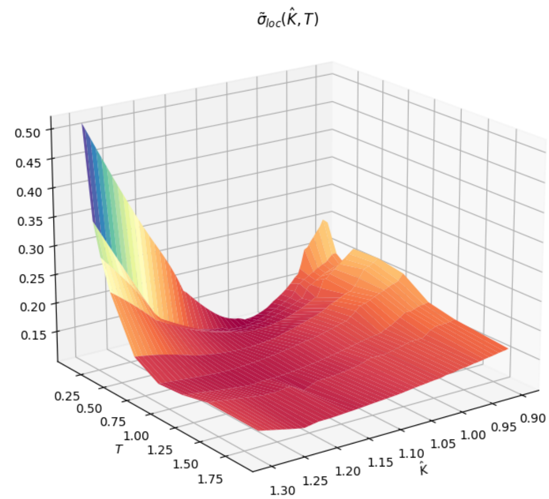


Figure 14: Local Volatility Surface

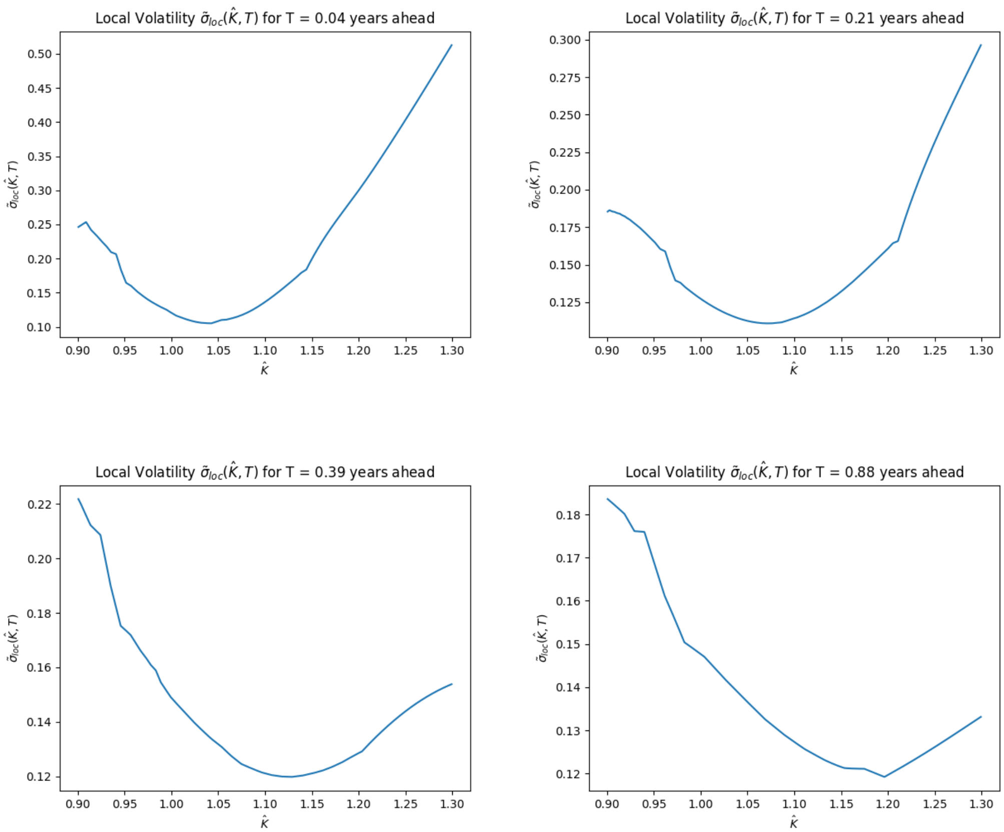


Figure 15: Local Volatilities for individual future times

4.5 Pricing Digital Options with Risk-Neutral Densities

In order to elaborate on the experiments using the Risk-Neutral Density for pricing contracts with general payoffs, the digital option was chosen given the availability of its closed-form prices in the Black-Scholes model, which serve as control experiments against our pricing with quadrature and RND. In both cases, local volatilities obtained from the model volatility surface were used as inputs to compute prices for digital European call options.

Figure 16: **Ground truth vs. density quadrature pricing For Digital asset-or-nothing Call Options on the Eurostoxx50 index**

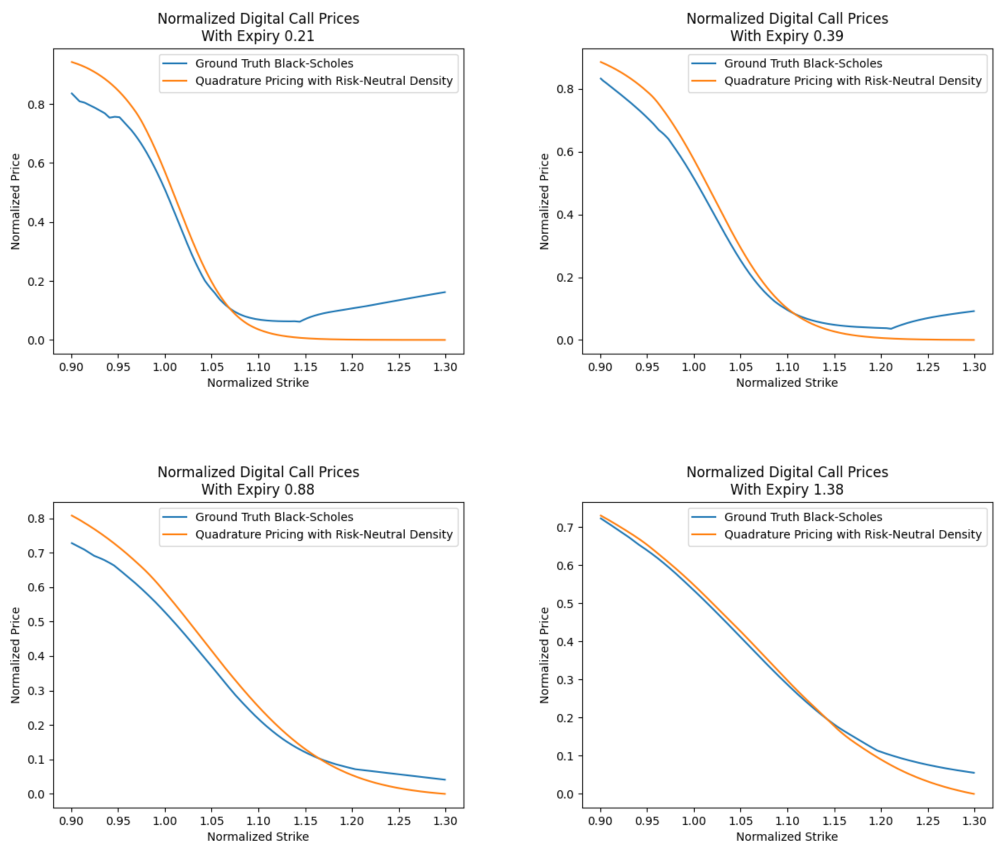


Table 5: Errors for quadrature pricing with RND of Digital European asset-or-nothing call options

Expiry	Mean Square Error
2.14e-01	8.64e-03
2.90e-01	2.76e-03
3.86e-01	2.95e-03
6.36e-01	4.49e-03
8.85e-01	2.16e-03
1.13e+00	1.06e-03
1.38e+00	5.80e-04

The reason that fewer expiries are presented above stems from the central difference scheme in (23), which remove four expiries: two in the beginning and two in the end. Also, only Mean Square Error is examined here because the benchmark prices and model prices both stem from the dense numerical grid, so reporting on data quantity becomes irrelevant (the data grid is not used, nor is the ratio $\frac{N_{numerical}}{N_{data}}$ relevant for the used $N_{numerical} = 2000$ as above), and total error become increasingly uninformative for dense grids.

5 Discussion

5.1 Grid Sparsity

The first set of experiments revealed that AHA model prices are not only always provably fair, but AHA also seemingly tolerates an arbitrarily coarse discretization of Dupire's equation. The results tell us that interpolating market prices with grids much more sparse than those used in the original paper by Andreassen and Huge [3] preserved near-perfect accuracy, enabling significant computational savings for the practitioner. In the original paper, they partition a range of $[0.5131, 1.466]$ in normalized strike into 200 points, while in our grid we split the much wider interval $[0.366, 2.604]$ into 60 points, meaning our stepsize in normalized strike is approximately five times that of the original paper [3]. Further coarsening of the strike-axis to 40 uniform steps over the same interval for $\hat{K} = \frac{K}{F_t(T)}$ cause visible imperfections in the calibration for the first expiry. At the same time, we still saw that model prices and market prices are indistinguishable for the longer expiries. For all expiries, mean square pricing errors are still desirably very low, staying below $1.03e-08$ across the data set.

A possible explanation for the marginally worse estimates with sparse grid selection is that coarsening the discretization induces truncation errors in the solution to Dupire's equation, meaning that even if we had perfect estimates of the volatility proxies(parameters), we lose accuracy in the numerical solver of eq. (22)(in contrast to losses due to misfitting the parameters). This could be the cause of our observed(small) losses in accuracy when adopting very sparse grids of merely 40 interpolation points.

While this behavior is expected and in line with basic facts about interpolating data, it raises the question of what a practitioner seeks to obtain from local volatility estimates. The literature voices divided opinions on what is desirable from such estimates. On the one hand, any local volatility surface that

1. Is void of arbitrage
2. Essentially matches the price data

is "perfect" in the sense that two such surfaces are equally valuable for any trading desk with access to only the above price data. This is one interpretation of the stance taken in the original paper by Andreassen and Huge [3]. Yet, acclaimed practitioners posit how sparse time steps in AHA impose truncation errors that hurt what they call "model accuracy" by flawed estimates of time-derivatives in the implicit finite difference scheme [2] [19]. The only plausible interpretation for such notions of "accuracy" is the degree to which the intermittent points between the data grid adequately solve Dupire's equation. This in turn implies that assumptions are made about the presence of an underlying "ground truth" volatility surface that is revealed through quotes emerging on the exchanges. To examine such claims, we can put ourselves in the shoes of a party or counterparty exposed to model risk[15] in a position long/short an exotic derivative, arising due to mere linear interpolation between the perfectly matching data grid points.

Abasto et.al [2], propose refining the time steps compared to the data grid in an attempt to overcome this concern. Yet, it is unclear what prices they optimize against with missing price data at the "new" expiries, in light of their stated use of the standard least squares loss. Any similar modeling attempt that adds more parameters(proxy volatilities) than the number of data points in an attempt to generalize better at intermittent points in either strike or maturity, we argue cannot capture any added information about the call surface (or equivalently, local volatility surface). Consequently, we argue that the model risk born by

[15]The higher order risk of the pricing model being mis-specified

both sides of the trading position as a consequence of sparse numerical grids is symmetric, and both the sell side and buy side cannot glean further insights into a ground truth volatility surface by refining the numerical grid, as was attempted by said researchers. However, the matter remains up for debate in the forthcoming research literature. Meanwhile, it should be noted that of course, a denser *data* price grid shrinks the set of arbitrage-free call surfaces, which reduces the uncertainty and model risk for both parties. AHA captures all the information available from such data sets by matching the number of parameters with the number of prices. We leave the matter open for researchers to debate.

5.2 Boundary Conditions

While sparse numerical grids caused essentially no damage to the quality of the model prices, removing the strike buffer around the boundaries caused concerns that were much more severe. A look at model vs. true prices for long maturities (1.8 and 2.8 years in Figure 5) reveals that the call surface obtained without buffering is useless in any real-world production environment: some AHA-prices grossly mismatch the market data, and subsequently likely expose model users of exploitable mispricings of instruments. Yet, results for short expiries appear satisfactory, which poses a whole new research question of why this is the case, i.e. why longer maturities are more vulnerable to this problem. We eliminated the hypothesis that grid sparsity was the root cause of model collapse at the boundary, by running a control experiment with an extremely dense grid of $10,000$ points, which gave us the same(inadequate) results.

A plausible cause for our results is that the boundary condition (39) are poorly reflected at the boundaries of the data grid, which brings about the need for buffering the strike boundaries. This is reflected in the first and last rows of the discretization matrix A, and wider boundaries then, hypothetically, support the boundary condition in the sense that those strikes are truly extreme, something in turn reflected by zero convexity (probability of asset approaching strike is close to zero). Logically, this makes sense, since the quotes listed by DBE that appear in our data would reasonably be used by hedgers, meaning that the market assigns non-zero probabilities to passing these levels at maturity. In contrast, the numerical grid needs zero such probability (zero convexity), and the neccessity of strike buffering follows.

5.3 Risk-Neutral Densities

Several discoveries were made when retrieving the Risk-Neutral Density(RND). We first comment on the interplay between the grid and and the RND. Above, we noticed very satisfactory price estimates in sparse grids. These did not transfer over to satisfactory RND estimates, but rather our RND estimator demands a much more dense numerical grid.

Despite seemingly correct price estimates from AHA model outputs, a particular inconsistency repeatedly emerged at the left part of the RND plots. Upon closer examination we found that the place where the flawed RND's assume abnormally large values lies at the point where the first market-strikes appear (the point where the left-side buffered strikes discontinue). This result offers several interpretations, one of which is that this is caused due to the optimization agent lacking information about market prices(and volatility) leftwards of the lowest strike. A reasonable consequence of this is then that calibrates on the left end, left of the data even, are less consistent with hypothetical market conditions, which would in turn say something about the implied density of underlying stocks. Subsequently, estimates of RND's cannot adequately infer implied densities around these points, which is what we see in the experiment results. A counterargument to this analysis is that this implies that poor density estimates would appear in the entire left region of the strike axis, not only the point where padded strikes transition into market strikes, which is a prediction that is inconsistent with our findings. Moreover, our analysis offers no explanation as to why this problem grows in maturity (the first four-or-so maturities does not suffer from this at all). A loose hypothesis of this is that these errors propagate in the AHA algorithm (prices from the prior maturity are used as inputs to the upcoming maturity).

A compelling strategy to enhance the smoothness of the call surface, and consequently, the RND, is to interpolate with higher order splines to maintain differentiability for the whole surface, in particular the knots at the grid nodes. Nevertheless, this would impose arbitrage into the call-surface [14], making the model useless for pricing. Researchers have expressed varying degrees of perceived importance for call surface regularity. While LeFloc'h places great emphasis on this [15] [14], the original paper as well as Rasmussen [3] [23] celebrates the coarse interpolation for its ability to preserve arbitrage freedom, with little mentions of differentiability. Others have suggested second order differentiability as the desirable criterion [25].

We conclude that these criteria depends on the use case for the model: call prices are highly satisfactory for the linear interpolation of AHA, while using RND for pricing or other modeling purposes such as analysis of the stock densities themselves, demands greater regularity conditions on the call surface from which it is derived. If the practitioner is pursuing only the latter without any desire to price and sell derivatives to exposed hedgers, which demand provably arbitrage-free prices, we invite the practitioner to leverage cubic splines to smoothen the density. One can see the resulting improvements upon doing so in the plots of figure (13).

5.4 Local Volatility retrieval

Following in the footsteps of Rasmussen [24], we attempted to invert Dupire's equation (22) for small timesteps between the market expiries, using the scheme in (23). The model call prices for the intermediate steps are in this scheme obtained by solving

$$\mathbf{A}\hat{\boldsymbol{c}}_{q+u} = \hat{\boldsymbol{c}}_q \ , \tag{49}$$

for $u \in \{1, 2, 3, \ldots, N_{smallSteps}\}$ where $N_{smallSteps}$ is the number of intermediate small time steps in between market expiries, chosen by the model user in an attempt to achieve a more dense local volatility surface between possibly coarse market expiries. Note that the proxy volatilities for the most recently fitted parameters (those for expiry q in this case) are used to populate the A-matrix in (49). The above can be interpreted as interpolating a denser expiry grid if such needs arise for the practitioner. Using this technique, we were not able to obtain the desired results of retrieving a representative local volatility surface. The experiment thus forms a failure to reproduce previously reported results, assuming our implementation was consistent with that described in [24]. Leaving the possibility for computer bugs aside, we speculate that the cause of this result could be how stepsizes in strike and maturity interplay, since our errors lessened(but persisted) in this method when using very dense strike axes, and only two extra time points between each market expiry. This phenomenon appears theoretically for solving Dupire's equation, which is the heat equation subject to the renowned stability criterion

$$\alpha \frac{\Delta T}{\Delta \hat{K}^2} \underset{for\ stability}{\leq} \frac{1}{2} \ , \tag{50}$$

where the diffusion coefficient α in the case of Dupire's equation is $\alpha_{Dupire} := \frac{1}{2}\hat{\sigma}^2(\hat{K}, T)\hat{K}^2$. When plotting call prices from the dense timestep interpolation, we saw adequate results, implying instability during differentiation as the root cause (by the logic of eliminating causes). Robust dense interpolation along the maturity axis would allow practitioners even more flexibility to price exotic claims, which is why resolving this is important, and we encourage future studies to find ways of reliably retrieving dense Local Volatility surfaces, not only strike-wise. When using the original market maturities, numerical differentiation became stable and we were able to collect a Local Volatility surface in line with quantities predicted by market data and theory, as can be seen in figures 15 and 16.

5.5 Pricing Digital Options with Risk-Neutral Densities

The estimated Local Volatility was used to price digital call options with payoff $\phi(S_T) = S_T \cdot \mathbb{1}_{S_T > K}$, and forms an example of how practitioners can use the model output from AHA, or other Local Volatility models, as inputs to other models in real-world scenarios. With solely market-implied volatility, we would have a scarce set of available strikes for pricing. The denser model surface from AHA, which also encodes a more realistic non-Black-Scholes Model that accommodates continuous-time volatility dynamics, allowed us to price digital options of many more strikes. This allows more flexibility to service buy-side option users. Moreover, we did this using our quadrature-density approach: a method either rarely explored or absent in previous research. Comparing resulting prices with analytically priced Black-Scholes ground-truth benchmarks using the same Local Volatilities as inputs, we found that our approach has some ability to accurately price options with non-standard payoffs. Meanwhile, some damage to price accuracy was done when moving from merely repricing European calls to pricing digital calls. A possible research endeavor for future studies is examining the efficacy of the quadrature approach for a wider variety of payoffs.

5.6 Future Research

The most pressing concern appearing in our result pertain to the model-prices extrapolated to buffered strikes beyond the data grid. While provably void of arbitrage and visually seemingly reasonable, the Risk-Neutral Densities revealed that our model outputs for our specific data assigned unrealistic probabilities deep in the money right at the far left end of the strike axis. Under the Local Volatility modeling assumption, the model call-surface, if accurate, holds predictive power of the probability density of the underlying asset price by virtue of eq. (10). Its apparent failure to do so at the left strike boundary raises important research questions. Possibly, unstable numerical differentiation is responsible for the inadequate RND's. Yet, our control-experiments using extremely dense grids suggests this is not the case, as the problem persisted. This leaves us to question what really occurs at the boundaries in AHA, and if there is an inherent property of the algorithm that causes this behaviour. The matter is especially peculiar given that this phenomena seems to propagate to solutions for longer maturities. We leave studying the underlying cause of these results for future researchers to explore.

6 Conclusion

In this work, a seemingly novel approach for options pricing using local volatility, Risk Neutral Density, and quadrature integration was proposed and examined. While able to adequately reproduce the prices of the original Vanilla European call options used to calibrate the model in the first place, the pricing scheme suffered from clear errors when used to price digital European call options. We interpret this result as the method maintaining relevance for future study, where advances in RND estimation could also bring about improvements in quadrature pricing as a downstream benefit.

Moreover, our detailed study of grid selection and mitigation of boundary conditions resulted in the finding that AHA is very robust to even extremely sparse grids, where the number of numerical grid nodes need not be much denser than the original data grid. Albeit previous studies celebrated this property of AHA, this study forms a more systematic examination of the matter whereas prior research chose grids more arbitrarily. Meanwhile, we acknowledge that there is a competing interest for model users since one of the very purposes of Local Volatility calibration is to obtain Local Volatility surfaces on an arbitrarily dense grid. On the other hand, sparse numerical grids offer faster calibration times, which is also an attractive feature in many real-world scenarios. Since our results imply that AHA is robust under almost any grid selection, we conclude that the practitioner has the freedom to chose grid depending on their needs, with reliable accuracy characteristics.

Finally, we acknowledge that grid selection had very different effects on the different parts of our study. For obtaining call-surfaces, AHA is very robust to sparse grids, as has been mentioned. Nevertheless, for obtaining RND's and also Local Volatility surfaces from call surfaces, grid selection immediately rose in importance. Given that both rely on numerical differentiation to obtain these quantities, a natural interpretation is that numerical differentiation is unstable particularly for call surfaces obtained from Local Volatility models, which is something that has been mentioned in prior research [1]. Grid selection in the AHA algorithm varies significantly depending on the needs of the practitioner.

References

[1] *6all.dvi*. https://people.maths.ox.ac.uk/trefethen/6all.pdf. (Accessed on 10/23/2024).

[2] Damian Abasto, Bernhard Hientzsch, and Mark Kust. "Monte Carlo pricing with local volatility grids". In: *Available at SSRN 2252916* (2013).

[3] Jesper Andreasen and Brian Huge. "Volatility interpolation". In: *Risk* 24.3 (2011), p. 76.

[4] Tomas Björk. "Arbitrage Theory in Continuous Time". In: (2019).

[5] Douglas T Breeden and Robert H Litzenberger. "Prices of state-contingent claims implicit in option prices". In: *Journal of business* (1978), pp. 621–651.

[6] Peter Carr and Dilip B Madan. "A note on sufficient conditions for no arbitrage". In: *Finance Research Letters* 2.3 (2005), pp. 125–130.

[7] Samuel N Cohen, Christoph Reisinger, and Sheng Wang. "Detecting and repairing arbitrage in traded option prices". In: *Applied Mathematical Finance* 27.5 (2020), pp. 345–373.

[8] *Data Shop Deutsche Börse AG - Your road to historical data - Data Shop Historical Data*. https://a7-dataplatform.deutsche-boerse.com/. (Accessed on 10/08/2024).

[9] *Deployments · fortran-lang/minpack*. https://github.com/fortran-lang/minpack/deployments. (Accessed on 11/20/2024).

[10] Emanuel Derman and Michael B Miller. *The volatility smile*. John Wiley & Sons, 2016.

[11] Fabrice Deschâtres. "Convex Volatility Interpolation". In: *Available at SSRN 4831218* (June 2024). URL: https://ssrn.com/abstract=4831218.

[12] Rick Durrett. *Probability: theory and examples*. Vol. 49. Cambridge university press, 2019.

[13] Stephen Figlewski. "Risk-neutral densities: A review". In: *Annual Review of Financial Economics* 10.1 (2018), pp. 329–359.

[14] Fabien Le Floc'h. "The Quadratic Local Variance Gamma Model: an arbitrage-free interpolation of class $C^3 foroptionprices$". In: *arXiv preprint arXiv:2305.13791* (2023).

[15] Le Floc'h et al. "Arbitrages in the Volatility Surface Interpolation and Extrapolation". In: *Available at SSRN 2175001* (2012).

[16] J Gatheral. *The volatility surface: A Practitioner's Guide*. John Wiley and Sons, Inc, 2011.

[17] Paul Glasserman. *Monte Carlo methods in financial engineering*. Vol. 53. Springer, 2004.

[18] Geoffrey Grimmett and David Stirzaker. *Probability and random processes*. Oxford university press, 2020.

[19] Alex Lipton and Artur Sepp. “Filling the gaps”. In: *Risk Magazine, October* (2011), pp. 66–71.

[20] *Local and Stochastic Volatility.* http://www.maths.lse.ac.uk/Personal/jruf/teaching/1507_StochVolSlides.pdf. (Accessed on 10/27/2024).

[21] Jorge J Moré, Burton S Garbow, and Kenneth E Hillstrom. *User guide for MINPACK-1.* Tech. rep. CM-P00068642, 1980.

[22] Bernt Oksendal. *Stochastic differential equations: an introduction with applications.* Springer Science & Business Media, 2013.

[23] L. Rasmussen. *Calibrating the local volatility model.* Københavns Universitet, Institut for Matematiske Fag, 2012. URL: https://books.google.se/books?id=OfM6zQEACAAJ.

[24] Lykke Rasmussen. “Computational Finance: On the Search for Performance”. PhD thesis. University of Copenhagen, Faculty of Science, Department of Mathematical . . ., 2016.

[25] Martin Tegner and Stephen Roberts. “A Bayesian take on option pricing with Gaussian processes”. In: *arXiv preprint arXiv:2112.03718* (2021).

TRITA – SCI-GRU 2024:467
Stockholm, Sweden 2025

www.kth.se