



UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2020:16

A Comparison between Approximations of Option Pricing Models and Risk-Neutral Densities using Hermite Polynomials

Nathaniel Ahy

Examensarbete i matematik, 30 hp
Handledare: Benny Avelin
Examinator: Erik Ekström
Juni 2020



Department of Mathematics
Uppsala University

Contents

1	Introduction	1
1.1	Background of Hermite Polynomials	1
1.2	Financial Derivatives and Stochastic Volatility	2
1.3	Hermite Polynomials in Finance	5
1.4	Problem Formulation	7
1.5	Methodology	7
1.6	Outline	8
2	Theory	10
2.1	Heston Model	10
2.1.1	The Fundamental Theorem of Finance	11
2.2	Gram-Charlier Expansions	13
2.2.1	Cumulants and an Explicit Formula for Computing them	13
2.2.2	Gram-Charlier Expansions for the Heston Model	14
2.3	Gram-Charlier Expansion Option Pricing Formula	14
2.4	Necula, Drimus and Farkas Risk-Neutral Measure	15
2.4.1	A closed form formula for the coefficients for the risk- neutral density	17
2.4.2	Implementation of Closed Form Formula	19
3	Option Pricing	21
3.1	Estimating Coefficients through quadratic programming	21
3.1.1	The Active Set Algorithm	25
3.2	Fitting the NDF log mean and standard deviation to a Model	27
4	Model Fitting Results	28
4.1	Computing the Cumulants	28
4.2	Model Errors	29

5	Option Data	33
5.1	Fitting the GC Parameters to the Data	33
6	Data Fitting Results	36
6.1	NDF Errors	36
6.2	GC errors	40
6.3	Model Complexity	43
7	Conclusion	45
7.1	Conclusion	45
7.2	Further Studies	46

Abstract

In this paper we discuss and display the broad applications of Hermite polynomials in mathematical finance. We will primarily focus on two of the applications thereof. Firstly, we consider how Hermite polynomials can be used to approximate the Heston stochastic volatility model [1], which can be seen in Cheng [2]. Secondly, we examine the manner in which Necula, Drimus, and Farkas [3] showed how these polynomials can be used to construct a risk-neutral probability density which serves as a generalization of the Black–Scholes formula [4]. To evaluate the performance of these models we compared how well they fit to the Heston model and thereafter how they, along with their more well known counterparts (the Black & Scholes and Heston models), fit to Amazon call option data using the trust region reflective nonlinear least squares and active-set quadratic programming methods. We analyze how the model accuracy varies using more moments in the Gram-Charlier approximation of Heston option prices; an approximation derived from Hermite polynomials whilst taking numerical stability into account. We also observe how varying the degree of the Hermite polynomial in Necula et. al.’s risk-neutral density effects the respective model’s ability to fit to option prices in various settings. We conclude that the Gauss-Hermite risk-neutral density introduced in Necula, Drimus, and Farkas significantly outperforms the Gram-Charlier method in the Heston setting in addition to the Black–Scholes model and has a distinguishable upper hand on the pure Heston model.

Chapter 1

Introduction

1.1 Background of Hermite Polynomials

Anders Hald gives a rich description of the history behind Hermite polynomials in his paper published in 2000 [5]. According to Hald the emergence of Hermite polynomials began with Laplace in his 1810 work on the central limit theorem. Poisson and Bessel in 1829 and 1838 respectively worked in parallel of Laplace and made major contributions in deriving Gram-Charlier expansions. Independent of Laplace Chebyshev in 1859 was inspired by Fourier's representation of an arbitrary function as an infinite trigonometric series and decided to use polynomials orthogonal with respect to a weight function¹ for the sake of linear/polynomial regression. Upon choosing the weight function e^{-x^2} Chebyshev discovers that Hermite polynomials served as the best means for approximating a function. Hermite (1864) defines these polynomials as:

Definition 1. The physicist Hermite polynomial $H_n(x)$ takes on the following form:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = \left(2x - \frac{d}{dx}\right)^n \cdot 1,$$

where it can easily be verified that $H_0(x) = 1$, $H_1(x) = 2x$.

In the same paper Hermite discusses the properties of said polynomials along with the infinite series they entail in approximating functions and remarks that they belong to a wide series of expansions that give interpolation

¹The polynomials $P_n(x)$ and $P_m(x)$ are said to be orthogonal with respect to a weight function $\xi(x)$ if.f. $\int_{-\infty}^{\infty} P_n(x)P_m(x)\xi(x)dx = 0$.

formulas derived by the method of least squares. Gnedenko and Kolmogorov (1954) called these polynomials Chebyshev-Hermite polynomials presumably not knowing that they originated by Laplace which could serve as an explanation as to why they are most commonly referred to as Hermite polynomials now. Thiele and Gram were among the first known to apply these polynomials to market and demographic data. Thiele in 1873 and in 1879 Gram (Jogen Pedersen Gram of whom, along with Carl Charlier, the Gram-Charlier expansion was named after) introduced the concept of Hermite polynomials in finance as they believed that the normal distribution did not sufficiently capture the properties of economic data. Thiele in 1889 proposed the density

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2} \sum_{r=0}^{\infty} \frac{c_r H_r(x)}{r!},$$

however he argues that: "In the majority of its purely mathematical applications $e^{-\pi x^2}$ is preferable (to the Gauss distribution)[...]" . However, Thiele was far ahead of his time as Gram-Charlier expansions do not seem to have been commonly applied into finance until after Jarrow and Rudd (1982) [6] who used Gram-Charlier expansion to improve the Black and Scholes [4] option pricing model 100 years later.

1.2 Financial Derivatives and Stochastic Volatility

Derivative pricing is among the most popular topics in financial mathematics. It was brought into the spotlight in part by Black and Scholes (BS) [4] who used no-arbitrage arguments to price European options.

Definition 2. An option provides its owner with the right but not the obligation to receive a payment depending on some financial asset.

Definition 3. A European call option upon its expiration provides its owner with the right but not the obligation to purchase an asset for a strike price K on some agreed upon expiration date.

These two Chicago based professors convincingly illustrated that the drift of an underlying asset has no bearing on the derivative price as that would

otherwise allow arbitrage opportunities; a risk-free profit. One major shortcoming of the Black and Scholes model is that it assumes volatility is constant; when it is evident by many volatility measures such as the VIX-index that volatility is, in itself, random. In particular Black and Scholes assume that the asset $S(t)$ follows a geometric Brownian motion which is arguably the most famous diffusion process to describe an asset price process. The reason for this is that this model arguably, in the simplest possible manner, captures commonly observed properties in assets: they have a compounding effect but it is intuitive to assume that they follow some sort of stochastic process with a drift. If S has a mean log-return μ , and log volatility σ , then the geometric Brownian motion asset price process of S is described as:

$$dS_t = \mu S_t + \sigma S_t dW_t. \quad (1.1)$$

Many ways of adjusting the volatility in the BS setting have been considered, such as the constant elasticity of variance (CEV) model popularized by John Cox in an unpublished draft of an article in 1975 [7]. The purpose of this model is to capture some nuances of volatility of which the BS fails:

$$dS_t = \mu S_t dt + \sigma S_t^\gamma \cdot S_t dW_t.$$

This model can potentially increase the volatility, $\sigma(S) = \sigma S^\gamma$, as the asset price increases when $\gamma > 0$ or, as commonly observed in equity stocks, the volatility increases as the price decreases, when $\gamma < 0$.

Although, upon observing common measures of market volatility such as the VIX index one notices that volatility in itself does not seem so well-behaved. This spurred the birth of stochastic volatility models. Among the first attempts to capture this was the 1982 discrete time Autoregressive Conditional Heteroskedasticity (ARCH) model by Robert Engle [8]. This model uses previous error terms to describe the current error hence why it is "Autoregressive". Denoting the error as ϵ_t , a random variable as z_t which is traditionally normally distributed with variance t , and a time dependent standard deviation ν_t the $ARCH(q)$ model is written as:

$$\begin{cases} \epsilon_t = \nu_t z_t, \\ \nu_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2. \end{cases}$$

Bollerslev in 1986 [9] then derived a Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model, where ν_t is regressed by its own previous

realizations, that is, the $GARCH(p, q)$ model is written as:

$$\nu_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \nu_{t-i}^2.$$

The $GARCH(1, 1)$ model in continuous time is written as:

$$d\nu_t = \kappa[\theta - \nu_t]dt + \sigma\nu_t dB_t.$$

We have the components

- ν_t - variance of the underlying asset at time t .
- κ - mean reversion,
- θ - long term volatility,
- σ - volatility of volatility.

This model only differs from the 1993 Heston [1] variance process in the sense that the diffusion term is the square-root of the volatility:

$$d\nu_t = \kappa[\theta - \nu_t]dt + \sigma\sqrt{\nu_t}dB_t.$$

The Heston model assumes that the asset follows a geometric Brownian motion but with stochastic volatility $\nu(t)$ in the place of the constant volatility, σ , in equation (1.1):

$$dS_t = \mu S_t dt + \sqrt{\nu_t} S_t dW_t.$$

and the Wiener processes W and B have correlation ρ . Heston illustrates that his stochastic volatility equipped with a correlation coefficient between the aforementioned Wiener processes explains return skewness and kurtosis which in turn effects out-of money relative to in-the-money pricing. The reason for this is that without stochastic volatility in all finite time spans returns are assumed to be log-normally distributed; a distribution known to underestimate the likelihood of events far from the mean. Additionally Heston uses his model to illustrate strike price biases in the Black-Scholes model.

Another popular stochastic volatility model, the SABR model, was developed

by Patrick S. Hagan, Deep Kumar, Andrew S. Lesniewski, and Diana E. Woodward [10] in 2002. It is written as:

$$\begin{cases} dF_t = \sigma_t F_t^\beta dW_t, \\ d\sigma_t = \alpha \sigma_t dZ_t, \end{cases}$$

where F_t is the forward rate (the interest rate of a transaction that will occur in the future), σ_t is the volatility, α and β are constant and W and Z have correlation ρ , i.e. $E[dW_t dZ_t] = \rho dt$. Notice that if $\alpha = 0$ then the forward rate is equal to the diffusion term of the CEV model. This is understandable as this is the very component of an asset price which a forward contract hedges one against.

1.3 Hermite Polynomials in Finance

Cheng [2] in their master thesis displays a method of using Gram-Charlier (GC) expansions; a so called Gram-Charlier A series to price financial derivatives under both the Heston and Black-Scholes model. This was done by deriving an infinite series representation of a probability distribution using only the normal distribution function, the log-moments of the underlying random variable, and "probabilist" Hermite polynomials, $He_n(x)$.

Definition 4. The probabilists Hermite polynomial $He_n(x)$ is defined as follows

$$He_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2} = \left(x - \frac{d}{dx} \right)^n \cdot 1.$$

It is easily verified that $He_0(x) = 1$ and $He_1(x) = x$. This means that one only needs to compute the moments $(E[X^n])_{n=0}^\infty$ of a random variable X which follows a certain distribution $f(x)$ to find the Gram-Charlier A series representation of $f(x)$. Because of its reliance on the derivative operator Hermite polynomials can, in many cases, make integration easier which underpins some of the convenience of using these polynomials.

Jarrow and Rudd [6] argue that in many cases certain characteristics about the underlying asset's distribution may be known but the distribution may be too complex to integrate over, making it hard to compute any of the asset price's moments. To alleviate this they approximate the distribution using the first four moments of the asset similar to that done in Cheng's

paper pioneering the applications of Gram–Charlier (GC) expansions in finance. They then expand upon the Black–Scholes model by including these moments in addition to the Black–Scholes formula. They test their approach by viewing how well their model fits to jump-diffusion and constant elasticity of variance asset pricing approaches. Unfortunately, Jarrow and Rudd conduct no tests on how their model fits to option data. Corrado and Su [11] built on their work by conducting these tests on S&P 500 option prices. They infer the kurtosis and skewness through their implied values from option data in the same manner as one infers implied volatility.

Definition 5. The implied volatility, σ^* , of some option price given by some model with parameters \vec{p} and volatility σ , $C_{model}(\sigma; \vec{p})$, inferred from some option price from data, $C_{Data}(\vec{p})$, is the value σ^* that solves the (typically nonlinear) equation:

$$\sigma^* = \sigma : C_{model}(\sigma; \vec{p}) = C_{Data}(\vec{p}).$$

Corrado and Su extend the above methodology of finding the implied volatility to simultaneously find the skewness and kurtosis. The results of Corrado and Su display that options priced under the Black–Scholes model for various dates differed from the data on average in absolute terms of 0.42–0.62 USD the average of these values being 0.4725 USD. Jarrow and Rudd’s model, on average, differed from the data in the interval of 0.16–0.51 USD, averaging over the different time periods at 0.2858. Corrado and Su then made the point that skewness alone dropped the average difference over the time periods down to 0.3092USD, meaning that the kurtosis term in Jarrow and Rudd’s model played a small but not a negligible role. They then argue that including higher moments would both lead to unstable estimations and that the even numbered moments are highly correlated. Additionally, it seems that the higher moments will also play an inconsequential role in the option price. These are arguments we will bear in mind when conducting our research especially when we price options using Cheng’s GC expansion pricing approach.

Necula, Drimus, and Farkas (NDF) build on this large body of option pricing research using Hermite polynomials but they used a different type of polynomial than Cheng; the physicist Hermite polynomial of Definition 1.

In their paper a risk-neutral density for the underlying asset is derived using a Gauss-Hermite (GH) infinite series (i.e. with "physicist" Hermite

polynomials) relying solely on the assumption that the log-return risk-neutral measure is characterized by a certain mean, standard deviation, and expansion coefficients. The reason for NDF deciding to use a different type of Hermite polynomials is that Cramer [12] shows that using high order coefficients in the GC A series could cause worse approximations as these expansions may fail to converge for fat-tailed distributions as it requires its related probability density function to fall off faster than $e^{-x^2/4}$. For simplicity when we refer to Gram-Charlier A series we will use the abbreviation 'GC' and when we refer to Gauss-Hermite expansions we will use 'GH'. These individuals then derive ways of inferring the coefficients of said expansion using the orthogonality condition of Gauss-Hermite polynomials as well as a method for deriving them through the observed option prices. To understand Necula, Drimus, and Farkas' results we need the following definition.

Definition 6. The volatility adjusted moneyness, m of an option paying no dividend with strike K , implied volatility σ^* , asset price at the birth of the option S_t , expiring at time T , continuously compounded risk-free interest rate r , and implied volatility σ^* is defined as:

$$m = \frac{\ln(K/S_t e^{r(T-t)})}{\sigma^*}.$$

Necula, Drimus and Farkas test their model by examining how well it adapts to option data. This results in them concluding that their model performs the best in-sample and out of sample for options with volatility adjusted moneyness in the $[-7,3]$ range.

1.4 Problem Formulation

Our goal is to see how well NDF's model can fit models in different scenarios in comparison to the GC method Cheng shed light on. The reason why this is done is to illustrate the flexibility of Hermite polynomials as well as the different ways in which they can be applied.

1.5 Methodology

Our approach to this problem is described in numerous steps.

- Price call options given a set of parameters \vec{p} under the Heston model using a closed form solution $C_H(K_i; \vec{p})$ for strike K_i where $i \in \{1, 2, \dots, N\}$.
- Price options using the Gram Charlier expansion approximation of the Heston model based on the first m moments of the asset price with the same set of parameters as Heston $C_{GC}^m(K_i; \vec{p})$.
- Given a set of parameters in the NDF setting, \vec{p}_{NDF} , calibrate the coefficients $\vec{a} \stackrel{def}{=} (a_j)_{j=0}^n$ for the NDF asset return density $P_{NDF}(x; \vec{a}, \vec{p}_{NDF})$ that minimizes the difference between the NDF option price, $C_{NDF}(K_i; \vec{a}, \vec{p}_{NDF})$, and $C_H(K_i; \vec{p})$.
- Equipped with \vec{a} find \vec{p}_{NDF} that best fits the NDF prices to the Heston prices.
- Compute $C_{NDF}(K_i; \vec{a}, \vec{p}_{NDF})$ and compare its relative difference of $C_H(K_i; \vec{p})$ with the relative difference of $C_{GC}^m(K_i; \vec{p})$ both in-sample and out-of-sample.
- Repeat the above steps for different n and m .

We will thereafter see how these models along with their more popular counterparts (Heston and Black–Scholes) perform on fitting data both in-sample and out of sample. This will be determined through repeating the above steps but we will fit the Heston parameters \vec{p} and the Black–Scholes only free parameter; its volatility, σ , to the data. After repeating the above itemization with the aforementioned steps we compare all our models.

1.6 Outline

We outline this paper in the following manner. In Part 2 we will discuss the theory as well as all the necessary prerequisites to provide the reader with a concrete idea of what we will be doing. The theory includes the background and definitions of Gram-Charlier series, the Heston model and the NDF model. We will then show how these concepts link to the two key papers of this thesis : Necula, Drimus, & Farkas [3] and Cheng [2]. Thereafter in Section 3 we briefly discuss how we implement these models numerically. In Section 4 we discuss how well the Gram-Charlier approximation of the

Heston model and NDF model did in approximating and fitting to the Heston model respectively. Section 5 focuses on how we fit the models to actual Amazon option data followed by Section 6 discussing the consequential results. Finally, we conclude with a summary and proposal for further studies.

Chapter 2

Theory

2.1 Heston Model

Heston's model is a one-factor stochastic volatility model describing an asset's price dynamics. The asset price process $S(t)$ and the volatility process $\sqrt{\nu(t)}$ are described as follows:

$$\begin{cases} dS(t) = \mu S(t)dt + \sqrt{\nu(t)}S(t)dW_1(t) \\ d\sqrt{\nu(t)} = \alpha\sqrt{\nu(t)}dt + \beta dW_2(t), \end{cases}$$

where $E[dW_1(t)dW_2(t)] = \rho dt$. Since $\nu(t) = \sqrt{\nu(t)}^2$ by Itô's lemma

$$\begin{aligned} dv(t) &= 2\sqrt{\nu(t)}d\sqrt{\nu(t)} + (d\sqrt{\nu(t)})^2 \\ &= [\beta^2 - 2\alpha\nu(t)]dt + 2\beta\sqrt{\nu(t)}dW_2(t) \\ &\stackrel{def}{=} \kappa[\theta - \nu(t)]dt + \sigma\sqrt{\nu(t)}dW_2(t), \end{aligned}$$

where we made the definitions:

- mean reversion: $\kappa = -2\alpha$
- long-term volatility: $\theta = \beta^2/\kappa$
- volatility of volatility: $\sigma = 2\beta$.

2.1.1 The Fundamental Theorem of Finance

The goal of pricing financial derivatives is to ensure that no arbitrage opportunities exist. That is, we wish to remove the possibility of achieving a risk-free profit under this model. An arbitrage free price is determined through ensuring that the change of a combination of financial derivative(s) and their underlying asset(s) matches the change of an equivalent initial investment into a risk-free asset. To more rigorously show what we mean we will first discuss the risk-neutral measure.

Definition 7. An asset S paying no dividends is said to follow a risk-neutral measure Q if its expected value discounted by the risk-neutral rate, r is its current price. That is to say:

$$e^{-rt}E^Q[S(t)] = S(0).$$

As we compute an expected value of a stochastic process we only concern ourselves with its drift; we only have to adjust the drift of the asset-price process to adhere to this risk-neutral measure. Allowing $S(t)$ to denote the asset price process, and letting $W^Q(t)$ be a standard Wiener process under the risk-neutral measure we get the following dynamics for the asset in Heston's model:

$$dS(t) = rS(t)dt + \sqrt{\nu(t)}S(t)dW_1^Q(t),$$

where $dW_1^Q(t) = dW_1(t) + \frac{\mu-r}{\sqrt{\nu(t)}}dt$. Proceeding, we consider the following entities:

- $S = S(t)$ - Underlying asset
- $C = C(S, \nu, t)$ - Financial derivative which we intend to price
- $U = U(S, \nu, t)$ - An instrument for hedging the volatility of S .

Using the above assets we construct a deterministic portfolio, $\Pi = C - w_1S - w_2U$, by eliminating the stochastic differential terms; $d\nu$ and dS through arguments similar to those in Gatheral [13].

$$d\Pi = \left(\frac{\partial C}{\partial t} + \frac{1}{2}\nu S^2 \frac{\partial^2 C}{\partial S^2} + \rho\nu\sigma S \frac{\partial^2 C}{\partial S\partial\nu} + \frac{1}{2}\sigma^2\nu \frac{\partial^2 C}{\partial\nu^2} \right) dt$$

$$\begin{aligned}
& -w_2 \left(\frac{\partial U}{\partial t} + \frac{1}{2} \nu S^2 \frac{\partial^2 U}{\partial S^2} + \rho \nu \sigma S \frac{\partial^2 U}{\partial S \partial \nu} + \frac{1}{2} \sigma^2 \nu \frac{\partial U^2}{\partial \nu^2} \right) dt \\
& + \left(\frac{\partial C}{\partial S} - w_1 - w_2 \frac{\partial U}{\partial S} \right) dS \\
& + \left(\frac{\partial C}{\partial \nu} - w_2 \frac{\partial U}{\partial \nu} \right) d\nu.
\end{aligned}$$

We easily solve for w_2 with w_1 following immediately after:

$$\begin{cases} w_2 = \frac{\partial C}{\partial \nu} / \frac{\partial U}{\partial \nu} \\ w_1 = \frac{\partial C}{\partial S} - \frac{\partial U}{\partial S} \frac{\partial C}{\partial \nu} / \frac{\partial U}{\partial \nu}. \end{cases}$$

Since Π is a deterministic portfolio with initial cost $C - w_1 S - w_2 U$ to satisfy no arbitrage constraints we get that

$$d\Pi = r(C - w_1 S - w_2 U)dt. \quad (2.1)$$

Since if this equality does not hold then one could short sell one of the portfolios, purchase the one that increases at a higher rate and enjoy a risk-free profit without spending anything at the time of the purchase. We will now define the first term characterizing $d\Pi$ as A and the second term as B . We can rearrange (2.1) to associate the financial derivative C with the volatility hedging asset U :

$$\frac{A - rC + rS \frac{\partial C}{\partial S}}{\frac{\partial C}{\partial S}} = \frac{B - rU + rS \frac{\partial U}{\partial S}}{\frac{\partial U}{\partial \nu}}.$$

We define what Heston refers to as the market price of volatility risk, λ , as follows:

$$\lambda(S, \nu, t) \stackrel{\text{def}}{=} \frac{B - rU + rS \frac{\partial U}{\partial S}}{\frac{\partial U}{\partial \nu}} + \kappa[\theta - \nu].$$

By rearranging the equation this definition results in we now have the Heston PDE:

$$\frac{1}{2} \nu S^2 \frac{\partial^2 C}{\partial S^2} + \rho \sigma \nu S \frac{\partial^2 C}{\partial S \partial \nu} + \frac{1}{2} \sigma^2 \nu \frac{\partial^2 C}{\partial \nu^2} + rS \frac{\partial C}{\partial S} \quad (2.2)$$

$$+ \{\kappa[\theta - \nu(t)] - \lambda\} \frac{\partial C}{\partial \nu} - rC + \frac{\partial C}{\partial t} = 0. \quad (2.3)$$

Letting $\tau = T - t$ Heston solves this PDE by assuming the solution takes on the form:

$$C(S, \nu, t) = SP_1 - Ke^{-r\tau}P_2, \quad (2.4)$$

he proceeds by solving a system equations for P_1, P_2 entailed by this function form by guessing the functional form of the characteristic function $f = \exp\{A(T - t) + B(T - t)\nu + i\phi x\}$, $x = \ln(S)$, and $i^2 = -1$, which reduces these equations to a system of ODEs to be solved for A and B . This then leaves us with the values for P_j as:

$$P_j(x, \nu, T; \ln(K)) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{e^{-i\phi \ln(K)} f_j(x, \nu, T; \phi)}{i\phi} \right] d\phi, \quad j \in \{1, 2\},$$

where

$$\begin{cases} C(\tau; \phi) = r\phi i\tau + \frac{a}{\sigma^2} \left[(b_j - \rho\sigma\phi i + d)\tau - 2 \ln \left[\frac{1 - ge^{d\tau}}{1 - g} \right] \right] \\ D(\tau; \phi) = \frac{b_j - \rho\sigma\phi i + d}{\sigma^2} \left[\frac{1 - e^{d\tau}}{1 - ge^{d\tau}} \right] \\ g = \frac{b_j - \rho\sigma\phi i + d}{b_j - \rho\sigma\phi i + d} \\ d = \sqrt{(\rho\sigma\phi i - b_j)^2 - \sigma^2((-1)^{j-1}\phi i - \phi^2)}. \end{cases}$$

We now have a solid foundation to build on for comparing NDF and the GC approaches to option pricing under this model.

2.2 Gram-Charlier Expansions

2.2.1 Cumulants and an Explicit Formula for Computing them

A Gram-Charlier Expansion is a means of representing a probability distribution in the form of an infinite series. To do this we use the probabilist's Hermite polynomials from Definition 4 in the introduction as well as the following definition.

Definition 8. The n th cumulant, c_n of a random variable X is

$$c_n = \frac{d^n}{du^n} \ln(E[e^{uX}]) \Big|_{u=0}.$$

As Utzet et. al. show in their paper [14] the moment generating function of the log-asset price in the Heston setting $M_t(u) = E[e^{uX_t}]$, where $X_t = \ln(S_t)$, is computed as:

$$M_t(u) = \exp\{x_0 u\} \left(\frac{e^{(\kappa - \sigma \rho u)t/2}}{\cosh(P(u)t/2) + (\kappa - \sigma \rho u) \sinh(P(u)t/2)/P(u)} \right)^{2\kappa\theta/\sigma^2} \exp \left\{ -\nu_0 \frac{(u - u^2) \sinh(P(u)t/2)/P(u)}{\cosh(P(u)t/2) + (\kappa - \sigma \rho u) \sinh(P(u)t/2)/P(u)} \right\}.$$

We can compute $\frac{d^n}{du^n} \ln(M_t(u)) \Big|_{u=0}$ symbolically and we have our coefficients for our infinite series.

2.2.2 Gram–Charlier Expansions for the Heston Model

In Cheng’s master thesis [2] they use the cumulants to represent the characteristic function of X in the form of a Maclaurin series. Thereafter, to represent the distribution of X in terms of a Gram-Charlier A series they use properties of Hermite polynomials and an inverse Fourier transform. Through this derivation the Gram-Charlier A series of the distribution function f of X is written as follows:

$$f(x) = \sum_{n=0}^{\infty} \frac{q_n}{\sqrt{c_2}} H_n \left(\frac{x - c_1}{\sqrt{c_2}} \right) \phi \left(\frac{x - c_1}{\sqrt{c_2}} \right), \quad (2.5)$$

where $q_0 = 1$, $q_1 = q_2 = 0$, and

$$q_n = \sum_{m=1}^{\lfloor \frac{n}{3} \rfloor} \sum_{\substack{k_1, k_2, \dots, k_m \\ k_1 + k_2 + \dots + k_m = n}} \frac{c_{k_1} \dots c_{k_m}}{m! k_1! \dots k_m! \sqrt{c_2}^n}, \quad n \geq 3. \quad (2.6)$$

2.3 Gram-Charlier Expansion Option Pricing Formula

In Cheng’s paper [2] they let $e^{X_t} = e^{-rT} S_t$ where T is the time of expiration and r is the risk-free interest rate. Thereafter Cheng computes the the

expectation $E^Q[e^{ax}\mathbb{I}\{x \geq \alpha\}]$ using the Gram-Charlier representation of the respective distribution from (2.5) which gives the value:

$$E^Q[e^{ax}\mathbb{I}\{x \geq \alpha\}] = e^{ac_1} \sum_{n=0}^{\infty} q_n J_n \left(\frac{\alpha - c_1}{\sqrt{c_2}}, a\sqrt{c_2} \right), \quad (2.7)$$

where c_n is the n th cumulant of X_t ; $c_n = \ln E[X_t^n]$ and

$$J_n(x, a) = \begin{cases} e^{a^2/2} N(a - x), & n = 0, \\ aJ_{n-1}(x, a) + H_{n-1}(x)\phi(x)e^{ax}, & n = \{1, 2, \dots\}. \end{cases}$$

The above identities enable us to price a European call. We define $k = \ln K - rT$ where K is the strike price of the underlying and T (we are letting the initial time $t = 0$) is the time to expiration, then the call option price, C , is given by:

$$C = E[(e^{X_T} - e^k)^+] = E[e^{X_T}\mathbb{I}\{X_T \geq k\}] - e^k E[\mathbb{I}\{X_T \geq k\}].$$

By (2.7) we get the following theorem.

Theorem 1. *A European call, C_H , in the Heston setting priced with a Gram-Charlier A series representation of the asset price density, with strike K , and asset cumulants $(c_i)_{i=1}^{\infty}$ can be written in the form of the following infinite series:*

$$C_H = e^{c_1} \sum_{n=0}^{\infty} q_n J_n \left(\frac{k - c_1}{\sqrt{c_2}}, \sqrt{c_2} \right) - e^k \left[N \left(\frac{c_1 - k}{\sqrt{c_2}} \right) + \sum_{n=3}^{\infty} (-1)^{n-1} q_n H_{n-1} \left(\frac{c_1 - k}{\sqrt{c_2}} \right) \phi \left(\frac{c_1 - k}{\sqrt{c_2}} \right) \right]. \quad (2.8)$$

2.4 Necula, Drimus and Farkas Risk-Neutral Measure

As opposed to using the probabilist's Hermite polynomials (acquired through differentiating $e^{-x^2/2}$) Necula, Drimus, and Farkas (NDF) use physicist Hermite polynomials, $H_n(x)$ defined in the introduction, that is:

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2} = \left(2x - \frac{d}{dx}\right)^n \cdot 1.$$

The reason for this is due to Cramér [12], page 223, equation 17.6.6a showing that probabilist's polynomial expansions will not converge for tails that fall off slower than $e^{-x^2/4}$, whereas in NDF's paper it is claimed that a result from a paper written in German by Myller-Lebedeff (1907) [15] implies that expansions using physicist Hermite polynomials, Gauss-Hermite expansions, converge even for fat-tailed distributions. The risk-neutral density of NDF's paper, $p(x)$, representing the log-return of an asset, S , with log mean return and standard deviation μ and σ respectively is written as

$$p(x) = \frac{1}{\sigma} z\left(\frac{x - \mu}{\sigma}\right) \sum_{n=0}^{\infty} a_n H_n\left(\frac{x - \mu}{\sigma}\right) \quad (2.9)$$

where $z(x)$ is the standard Gaussian distribution.

Through the orthogonality conditions of "physicist" Hermite polynomials the coefficients can be found from

$$a_n = \frac{\sqrt{\pi}}{2^{n-1}n!} \int_{-\infty}^{\infty} z\left(\frac{x - \mu}{\sigma}\right) H_n\left(\frac{x - \mu}{\sigma}\right) p(x) dx. \quad (2.10)$$

NDF also propose methods for deriving a_n from option prices through integrating over a range of strike prices as well as a method using optimization, both of which will be discussed in the next part of this paper. These methods will be used to infer the coefficients from the option prices computed by the formula displayed in (2.4) of the previous section.

To price a European call option we define the following parameters:

- $S(t) = S_t$ - realized price of the option's underlying asset at time t ,
- r - risk-free interest rate,
- q - Dividend-yield of underlying.

With the above parameters they derive, through integrating the option pay-off over the risk-neutral density (2.9), that the price of a European call, $c(S_t, K, \mu, \sigma, \tau, r, q) \stackrel{def}{=} c$ can be found through the theorem below.

Theorem 2. Assume that the log-return risk-neutral measure for a given time horizon has the annual mean, μ , standard deviation, σ , and Gauss-Hermite expansion coefficients a_n for all $n \in \mathbb{N}$. Then the price of a European call, c , with strike K and maturity $t + \tau$ is given by:

$$c = S_t e^{-q\tau} \Pi_1 - K e^{-r\tau} \Pi_2, \quad (2.11)$$

where

$$\begin{cases} \Pi_1 = \exp \left\{ \left(\mu - (r - q) + \frac{\sigma^2}{2} \right) \tau \right\} \sum_{n=0}^{\infty} a_n I_n \\ \Pi_2 = \sum_{n=0}^{\infty} a_n J_n. \end{cases}$$

The I_n and J_n terms satisfy recursive formulas:

$$\begin{aligned} I_{n+1} &= 2z(-d_1)H_n(-d_2) + 2\sigma\sqrt{\tau}I_n + 2nI_{n-1}, & n \in \{1, 2, \dots\} \\ J_{n+1} &= 2z(-d_2)H_n(-d_2) + 2nJ_{n-1}, & n \in \{1, 2, \dots\}, \end{aligned}$$

where J has no relation to Cheng's formula and $I_0 = N(d_1)$, $I_1 = 2z(-d_1) + 2\sigma\sqrt{\tau}N(d_1)$, $J_0 = N(d_2)$, $J_1 = 2z(-d_2)$, finally

$$\begin{cases} d_1 = \frac{\log(S_t/K) + (\mu + \sigma^2)\tau}{\sigma\sqrt{\tau}} \\ d_2 = d_1 - \sigma\sqrt{\tau}. \end{cases}$$

It is also worth noting that in the case when $n = 0$, i.e. we only consider the first term of the infinite series, equation (2.11) nearly reduces to the familiar Black-Scholes formula (see the exponential term in Π_1 to notice the difference).

2.4.1 A closed form formula for the coefficients for the risk-neutral density

After computing the options prices from Cheng's work [2] we have two potential methods which we can rely on from NDF's paper. Necula Drimus and Farkas use the orthogonality conditions and apply change of variables to (2.10) giving us the integral representation of a_n :

$$a_n = \frac{\sqrt{\pi}}{2^{n-1}n!} \int_0^\infty G_n(S_{t+\tau}) p_{t+\tau}(S_{t+\tau}) dS_{t+\tau},$$

where $\tau = T - t$ is the time to expiration and $p_{t+\tau}(S_{t+\tau})$ is the terminal underlying asset price risk-neutral density. Moreover,

$$\begin{aligned} G_n(S) &= z\left(\frac{\ln(S/S_t) - \mu\tau}{\sigma\sqrt{\tau}}\right) H_n\left(\frac{\ln(S/S_t) - \mu\tau}{\sigma\sqrt{\tau}}\right) \\ &\stackrel{\text{def}}{=} z(d_2(S)) H_n(-d_2(S)), \end{aligned}$$

where we define $d_2(S) = \frac{\ln(S_t/S) + \mu\tau}{\sigma\sqrt{\tau}}$ and since $z(x)$ is the standard Gaussian distribution, by symmetry $z(-x) = z(x)$. We can apply the above identity to a result in Bakshi et. al [16] equation (3). We can therefore compute a_n using the theorem from NDF below.

Theorem 3. *The risk-neutral density coefficients $(a_i)_{i=0}^\infty$ characterizing the log-returns of an asset S with the continuum of European call and put prices $c(S_t, K, \tau)$ and $p(S_t, K, \tau)$ respectively written on time t with strike $K \in \mathbb{R}^+$ are computed with the following formula*

$$\begin{aligned} a_n &= \frac{\sqrt{\pi}}{2^{n-1}n!} \left\{ z(d_2(S_t e^{(r-q)\tau})) H_n(-d_2(S_t e^{(r-q)\tau})) + \right. \\ &\quad \left. e^{r\tau} \left[\int_{S_t e^{(r-q)\tau}}^\infty F_n(K) c(S_t, K, \tau) dK + \int_0^{S_t e^{(r-q)\tau}} F_n(K) p(S_t, K, \tau) dK \right] \right\}, \end{aligned} \quad (2.12)$$

where $F_n(K) = G_n''(K)$.

The term $F_n(\cdot)$ is given more explicitly by:

$$\begin{aligned} F_n(K) &= \frac{z(d_2)}{K^2 \sigma^2 \tau} \left[(d_2^2 - \sigma\sqrt{\tau}d_2 - 1) H_n(-d_2) + \right. \\ &\quad \left. (4nd_2 - 2n\sigma\sqrt{\tau}) H_{n-1}(-d_2) + 4n(n-1) H_{n-2}(-d_2) \right], \quad n \in \{2, 3, \dots\}. \end{aligned}$$

Since $F_n(K) = G_n''(K)$ which can be seen upon viewing equation (3) in Bakshi et. al. (2003). By computing $G_0''(K)$ and $G_1''(K)$ we immediately get the explicit values for F_0 and F_1 .

$$\begin{cases} F_0(K) = z(d_2(K)) [(d_2(K)d_2'(K))^2 - d_2'(K)^2 - d_2(K)d_2''(K)] \\ F_1(K) = z(d_2(K)) [-d_2(K)^3 d_2'(K)^2 + d_2(K)d_2'(K)^2 + 2d_2(K)d_2'(K)^2 d_2''(K) - d_2''(K)]. \end{cases}$$

2.4.2 Implementation of Closed Form Formula

All the ingredients of the closed form formula for our coefficients (2.12) can be found in a straightforward manner using identities discussed earlier in this paper. The only exceptions are the the European puts and calls $p(S_t, K, t)$ and $c(S_t, K, t)$ respectively. We get the European call terms $c(S_t, K, t)$ from the Heston formula (2.4). The put-values are found using the put-call parity stated in the following theorem.

Theorem 4. *The Put-Call Parity: A European put, $p(S_t, K, t)$ and call, $c(S_t, K, t)$ must satisfy the relationship*

$$p(S_t, K, t) + S_t = c(S_t, K, t) + K e^{-r(T-t)}$$

.

To see why the above theorem is true consider the portfolios on the RHS and LHS: purchase a put with strike K and the underlying asset, purchase a call with strike K and $K e^{r(T-t)}$ units of the risk-free asset. Both these portfolios have the same payoff at time expiration time T :

$$(K - S(T))^+ + S(T) = (S(T) - K)^+ + K.$$

Thus, to prevent arbitrage opportunities the put-call parity must hold. To get numerical values we use the values, similar to Section 6.5 in Cheng [2] $S_t = 100$, $t = 0$, $T = 1$, $r = 0.04$, $q = 0$, and we consider $K \in \{40, 40.001, \dots, 170\}$ as the put and call prices are sufficiently close to 0 in the lower and upper bounds of this discretized interval. This allows us to reduce the bounds of the integral from (2.12) to the bounds of our discretized interval for K providing us with a reasonable approximation for our coefficients. Thus, our numerical approximation of (2.12) is

$$a_n = \frac{\sqrt{\pi}}{2^{n-1}n!} \left\{ z \left(d_2(S_t e^{(r-q)\tau}) \right) H_n \left(-d_2(S_t e^{(r-q)\tau}) \right) + e^{r\tau} \left[\sum_{i: K_i = S_t e^{(r-q)\tau}}^{170} F_n(K_i) c(S_t, K_i, \tau) \Delta K + \sum_{i: K_i = S_t e^{(r-q)\tau}}^{40} F_n(K_i) p(S_t, K_i, \tau) \Delta K \right] \right\},$$

where $\Delta K = 0.001$, $K_i = 40 + 0.001i$, $F_n(170)c(100, 170, 0) \approx 0$, and $F_n(40)p(100, 40, 0) \approx 0$. Unfortunately, we attempted to price options fitting $(a_j)_{j=0}^n$ to the Heston model but achieved an in-sample mean relative

error of -4.7 meaning that it miscalculated option prices by -470%! Naturally, due to this undesirable outcome we will rely on other methods to find the α -coefficients in our result section.

Chapter 3

Option Pricing

3.1 Estimating Coefficients through quadratic programming

An alternative method is to compute our coefficients through quadratic programming. To do this we must first translate our problem into quadratic programming terms:

$$\min_a \frac{1}{2} a^\top L a + f^\top a \text{ subject to } \begin{cases} A \cdot x \leq b \\ A_{eq} \cdot x \leq b_{eq}, \end{cases} \quad (3.1)$$

where a is our vector of coefficients and we will derive L and f . Clearly, with the a -coefficients of the Gauss-Hermite (GH) distribution we want to minimize the difference between the the corresponding call option under a certain method, $C(K_k)$ ¹ and the N NDF European call option prices given the log-return mean and volatility μ and σ respectively², $(C_{NDF}(K_k); a, \mu, \sigma)_{k=0}^N$, having the N strikes $(K_k)_{k=0}^N$ and . This problem translates to

$$\min_a \sum_{k=0}^N (C_{NDF}(K_k; a, \mu, \sigma) - C(K_k))^2 \quad (3.2)$$

We will also use the following shorthand notations

¹In this case we will be using the Heston model itself.

²These parameters are guessed initially and estimated after the a -coefficients are found.

- $S_t \exp\{(\mu - (r - q) - \sigma^2/2)\tau\} \exp(-q\tau) \stackrel{def}{=} \tilde{S}$
- $K_k e^{-r\tau} \stackrel{def}{=} \tilde{K}_k$
- $n + 1$ - the number of a -coefficients
- $I_{ik} - I_i$ for an NDF option with strike K_k
- $J_{ik} - J_i$ for an NDF option with strike K_k ,

recall that I and J are the terms involved in (2.11). Applying (2.11) to (3.2) and expanding the expression gives us the following problem,

$$\min_a \sum_{k=0}^N \left[\tilde{S}^2 \sum_{j=0}^n \sum_{i=0}^n a_i a_j I_{ik} I_{jk} - 2\tilde{S}\tilde{K}_k \sum_{j=0}^n \sum_{i=0}^n a_i a_j I_{ik} J_{jk} \right. \quad (3.3)$$

$$\left. + \tilde{K}_k^2 \sum_{j=1}^n \sum_{i=0}^n a_i a_j J_{ik} J_{jk} - 2 \left(\tilde{S} \sum_{i=0}^n a_i I_{ik} - \tilde{K}_k \sum_{i=0}^n a_i J_{ik} \right) C(K_k) + C^2(K_k) \right] \quad (3.4)$$

The sums above have independent indices making them interchangeable, we use the following notation:

- $\bar{I}_i \stackrel{def}{=} \sum_{k=0}^N I_{ik}$
- $\bar{J}_i \stackrel{def}{=} \sum_{k=0}^N J_{ik}$
- $\overline{I_i J_j} \stackrel{def}{=} \sum_{k=0}^N I_{ik} J_{jk},$

that is, a value $\bar{\ell}$ is ℓ summed over its k dimension, the notation for the remaining summations is trivial. This notation entails the following problem:

$$\min_a \left[\tilde{S}^2 \sum_{j=0}^n \sum_{i=0}^n a_i a_j \bar{I}_i \bar{I}_j - 2\tilde{S} \sum_{j=0}^n \sum_{i=0}^n a_i a_j \overline{\tilde{K} I_i J_j} \right. \quad (3.5)$$

$$\left. + \sum_{j=1}^n \sum_{i=1}^n a_i a_j \overline{\tilde{K}^2 J_i J_j} - 2 \left(\tilde{S} \sum_{i=0}^n a_i \overline{C(K) I_i} - \sum_{i=0}^n a_i \overline{\tilde{K} J_i C(K)} \right) + \overline{C^2(K)} \right] \quad (3.6)$$

To formulate this problem in terms of (3.1) we make the following matrix definitions to scale all the terms with their corresponding a -coefficients.

$$L = 2 \begin{pmatrix} \tilde{S}^2 \overline{I_0 I_0} - 2\tilde{S} \tilde{K} \overline{I_0 J_0} + \overline{K^2 J_0 J_0} & \dots & \tilde{S}^2 \overline{I_0 I_n} - 2\tilde{S}^2 \overline{K I_0 J_n} + \overline{K^2 J_0 J_n} & 0 \\ \vdots & \ddots & \vdots & \vdots \\ \tilde{S}^2 \overline{I_n I_0} - 2\tilde{S} \tilde{K} \overline{I_n J_0} + \overline{K^2 J_n J_0} & \dots & \tilde{S}^2 \overline{I_n I_n} - 2\tilde{S}^2 \overline{K I_n J_n} + \overline{K^2 J_n J_n} & 0 \\ 0 & \dots & \dots & 0 \end{pmatrix} \quad (3.7)$$

$$f^\top = -2 \begin{pmatrix} \tilde{S} \overline{I_0 C(K)} - \tilde{K} \overline{J_0 C(K)} & \dots & \tilde{S} \overline{I_n C(K)} - \tilde{K} \overline{J_n C(K)} & -\frac{1}{2} \overline{C^2(K)} \end{pmatrix}. \quad (3.8)$$

Defining

$$a^\top = (a_0 \quad \dots \quad a_n \quad 1), \quad (3.9)$$

one can easily verify that:

$$\min_a \frac{1}{2} a^\top L a + f^\top a = \min_a \sum_{k=0}^N (C_{NDF}(K_j) - C(K_j))^2.$$

We now have our objective function in place, the only constraints that need to be satisfied are the unitary mass constraint of the probability density:

$$p(x) = \frac{1}{\sigma} z \left(\frac{x - \mu}{\sigma} \right) \sum_{j=0}^n a_j H_j \left(\frac{x - \mu}{\sigma} \right),$$

the density is everywhere positive, $p(x) \geq 0$ for all x , and that the final element of the vector a , stated in (3.9) is 1 for the sake of incorporating the Gram-Charlier option price in our computation as can be seen when evaluating $f^\top a$. The first task is simply done by verifying that $\sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} a_{2j} \frac{(2j)!}{j!} = 1$ by a formula from NDF [3]. To ensure that the density is everywhere positive we define n as the degree of the polynomial in $p(x)$ and consider a grid $x \in [-3n, 3n]$. Thereafter we consider the discretization $-3n = x_0 < x_1 < \dots < x_M = 3n$. Thus, we will verify that $p(x_j) \geq 0$ for all $j \in \{0, 1, \dots, M\}$ thereby ensuring that our density is roughly everywhere positive. These three constraints correspond to the following matrix relationships:

$$\begin{cases} A \cdot a \leq b, \\ A_{eq} \cdot a = b_{eq}, \end{cases}$$

where A is $(M+1) \times (n+1)$ with the additional term in the column space being explained by (3.9):

$$A = - \begin{pmatrix} \frac{1}{\sigma} z \left(\frac{x_0 - \mu}{\sigma} \right) H_0 \left(\frac{x_0 - \mu}{\sigma} \right) & \dots & \frac{1}{\sigma} z \left(\frac{x_0 - \mu}{\sigma} \right) H_n \left(\frac{x_0 - \mu}{\sigma} \right) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sigma} z \left(\frac{x_M - \mu}{\sigma} \right) H_0 \left(\frac{x_M - \mu}{\sigma} \right) & \dots & \frac{1}{\sigma} z \left(\frac{x_M - \mu}{\sigma} \right) H_n \left(\frac{x_M - \mu}{\sigma} \right) & 0 \end{pmatrix},$$

it is important to notice the negative sign in front of the above matrix. We also have that b is $(M+1) \times 1$,

$$b^\top = (0 \quad \dots \quad 0).$$

The matrix A_{eq} is $2 \times (n+1)$. The first row of this matrix is intended for satisfying the unitary mass constraint of the probability density by a formula mentioned in NDF's paper: $\sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} a_{2j} \frac{(2j)!}{j!} = 1$. The second row is intended to ensure that the last element of a is 1 (3.9). This thereby gives us that

$$A_{eq} = \begin{cases} \begin{pmatrix} \frac{0!}{0!} & 0 & \frac{(2 \cdot 2)!}{2!} & 0 & \dots & 0 & \frac{(2n)!}{n!} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}, & \text{for even polynomials,} \\ \begin{pmatrix} \frac{0!}{0!} & 0 & \frac{(2 \cdot 2)!}{2!} & 0 & \dots & \frac{(2n)!}{n!} & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}, & \text{for odd polynomials} \end{cases}$$

and b_{eq} is 2×1 :

$$b_{eq}^\top = (1 \quad 1).$$

We are now left with the optimization problem

$$\min_a a^\top \frac{1}{2} L a + f^\top a \text{ subject to } \begin{cases} A \cdot a \leq b, \\ A_{eq} \cdot a \leq b_{eq}. \end{cases} \quad (3.10)$$

The above optimization problem is finally solved through the MATLAB function *fmincon*. We will be using the active set algorithm to solve this problem.

3.1.1 The Active Set Algorithm

This algorithm is thoroughly discussed in literature such as Numerical Optimization, Nocedal and Wright (2006) [17], however, we will include a description here so the paper is self-contained. This method considers a so called 'working region' for iteration k , \mathcal{W}_k , in which all the equality constraints and a subset of the inequality constraints are considered as equalities. The function first checks whether there exists a solution to this subproblem otherwise it ignores all the constraints outside the working set and optimizes with respect to the parameters involved in the working set subject to the working set constraints. We want to reframe (3.10) to a smaller sub-problem. Defining $q(a) = a^\top \frac{1}{2}La + f^\top a$ then for the sub-problem of taking the optimal step, x_k , when we are at point a^k , can be stated as:

$$\min_x q(a^k + x) = \min_x \frac{1}{2}(a^k + x)^\top L(a^k + x) + f^\top (a^k + x) \quad (3.11)$$

$$= \min_x \frac{1}{2}x^\top Lx + f^\top x + 2\frac{1}{2}(a^k)^\top Lx + \xi_k \quad (3.12)$$

$$= \min_x \frac{1}{2}x^\top Lx + (f^\top + (a^k)^\top L)x + \xi_k, \quad (3.13)$$

where ξ_k is the remaining a^k terms. We will set $g_k^\top \stackrel{def}{=} (f^\top + (a^k)^\top L)$. We now want to minimize q subject to our working region, we can ignore the a^k component leaving us with

$$\min_x \frac{1}{2}x^\top Lx + g_k^\top x \quad (3.14)$$

$$\text{subject to } A_i^\top x = 0, \quad i \in \mathcal{W}_k. \quad (3.15)$$

We can translate the above problem to :

$$\begin{pmatrix} L & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} -x \\ \lambda \end{pmatrix} = \begin{pmatrix} g \\ h \end{pmatrix}, \quad (3.16)$$

where $h = Ax - b$ and λ is a vector of the Lagrange multipliers. Nocedal and Wright solve the above problem by symmetric indefinite factorization; a way of performing triangular factorization on the involved matrices and solving the involved system of linear algebraic equations. This can be done since we only have equality constraints; giving us the value x_k .

In the case when $x_k \neq 0$ regardless we have, by the constraint, $A_i^\top (a^k + \alpha x_k) =$

$A_i^\top a^k$; meaning if a^k was already a feasible point then $a^k + \alpha x_k$ is too for any α . If $a^k + x_k$ satisfies all of our constraints including those outside the working set we make the assignment $a^{k+1} = a^k + x_k$, which clearly reduces the objective function since $x_k = 0$ is already a feasible solution. Otherwise we update as: $a^{k+1} = a^k + \alpha_k x_k$ where α_k is the maximum value in the interval $[0, 1]$ satisfying our constraints. Nocedal and Wright show in their book that $\alpha_i < 1$ either if for some $i \notin \mathcal{W}_k$ we have not satisfied some constraint(s) or we have reached the optimum. It is therefore, in this case, necessary to update the working region by including some constraint(s) that is(are) prohibiting us from taking the optimal step; a blocking constraint. This is done until we have properly minimized q in this region. Otherwise we will keep using our current working region.

If $x_k = 0$ then we compute Lagrange multipliers λ_i satisfying

$$\sum_{i \in \mathcal{W}_k} A_i \lambda_i = g = Lx_k + f.$$

Nocedal and Wright show in Theorem 12.1 that the above equality will along with chosen step size α satisfy the first three Karush-Kuhn-Tucker (KKT) conditions; which ensure that we are at a stationary and primal feasible point, which means we have found a solution but not necessarily an optimal one. However, we have found an optimum if $\lambda \geq 0$; since this implies dual feasibility which is an upper bound of the optimal primal solution. Moreover, since we can use $(L + L^\top)/2$ instead of L to achieve the same minima arguments, a , and it is easily verified that this matrix is a positive semi-definite then our optimum is global by Theorem 16.4 of Nocedal and Wright. Otherwise we remove all the $w_j \in \mathcal{W}_k \cap \mathcal{I}$, \mathcal{I} being the set of inequality constraints, such that $\lambda_j < 0$ and solve an alternative optimization problem. This entire procedure explains how the active-set option in the *fmincon* algorithm works. As previously mentioned, for a more thorough description we refer to Nocedal and Wright (2006).

3.2 Fitting the NDF log mean and standard deviation to a Model

We guessed μ and σ based on previous optimization sessions³ giving us the coefficients $(a_j)_{j=0}^n$. After having found the a -coefficients we will find our mean log-return and standard deviation: μ and σ respectively through non-linear least squares as these terms both act on the exponents of our option pricing models. We let $\vec{p} \stackrel{def}{=} (\mu, \sigma)$ denote the vector of our parameters and we consider the optimization problem:

$$\min_{\vec{p}} \sum_{i=0}^n (C_{NDF}(K_i; \vec{p}) - C_{Data}(K_i))^2.$$

In the nonlinear least squares method we use, trust region-reflective least squares, we take a function $q(\vec{p})$ that approximates C_{NDF} for a small region (trust region) and minimize it. To find q we conduct a second order Taylor approximation of C in its current region. The problem translates to

$$\min_s \frac{1}{2} s^\top H s + s^\top g \quad \text{subject to } \|Ds\|_2^2 \leq \Delta, \quad (3.17)$$

where H is the Hessian, g is the gradient D is a diagonal scaling matrix and s is our parameter set within this current 'trust region'. To solve (3.17) ideally we reduce the space of s values to a two dimensional space S . The space S is spanned by s_1 and s_2 , with s_1 being the direction of the gradient and s_2 satisfying one of the two relations:

$$\begin{cases} Hs_2 = -g. \\ s_2^\top Hs_2 < 0, \end{cases}$$

if the function we wish to optimize f satisfies $f(x+s) < f(x)$ where x is our current point we update $x \leftarrow x + s$. Otherwise we shrink the trust region Δ and repeat. This entire process is then repeated until the parameters' difference between iterations falls below a pre-specified optimality tolerance. All of this is done by the MATLAB function *lsqnonlin*⁴.

³In the context of the Amazon option prices these values were $\mu = r + 0.01$ and $\sigma = 0.3$

⁴A slightly more thorough description of the trust-region reflective algorithm can be found here: <https://se.mathworks.com/help/optim/ug/least-squares-model-fitting-algorithms.htmlbroz0i4>

Chapter 4

Model Fitting Results

4.1 Computing the Cumulants

Recall that when we say the n th cumulant we refer to the n th moment of the log-asset price. We computed the cumulants symbolically through both MATLAB and Mathematica online by implementing and differentiating $M_t(u)$ described in Section 2.2.1 where the cumulants in MATLAB led to our GC option pricing method having a smaller relative error with regard to the Heston model. This could in part be attributed to mathematica's online platform perhaps having a higher round-off error than MATLAB. Regardless, for the remainder of this paper we will rely on our MATLAB cumulants. We will compute our cumulants using the same parameters as those used in Cheng's master thesis [2] p.66:

- initial volatility - $\nu_0 = 0.03$
- mean reversion - $\kappa = 0.15$
- volatility of volatility - $\sigma = 0.05$
- correlation between variance and asset's Wiener processes - $\rho = -0.55$
- time to expiration - $T = 1$
- annual log return of risk-free asset - $r = 0.04$
- initial asset price - $S_0 = 100$

- strike price for a European call option - $K \in \{40, 40.001, \dots, 170\}$,

where the set of K -values was based on the put and call options being virtually 0 on the respective lower and upper bound of the set. This yielded the first eight cumulants having the values:

i	c_i
1	4.589456320893087
2	0.031838851938017
3	-0.001261815533270
4	0.000115992103852
5	-0.000011185767429
6	0.000001414417922
7	-0.000000200913900
8	0.000000027458226

This clearly shows that the higher moments play a less significant role in the final option price than the lower moments, this will be made more clear in the next section.

4.2 Model Errors

To see how compatible NDF's and the Gram–Charlier¹ option pricing models are with the Heston model we will fit the former to the latter. We will consider an $m \in \{4, 6, \dots, 14\}$ degree polynomial for the Gauss–Hermite risk-neutral density² behind NDF's option pricing method and $n \in \{1, 2, \dots, 8\}$ coefficients (recall the q_n -terms derived in Section 2.2) for the GC option pricing formula. We make the following definitions.

- $C_{NDF}^m(K)$ - the option price in the NDF setting with an m th degree Gauss–Hermite polynomial as the risk-neutral density (2.9).

¹Recall that we abbreviate this to GC, the infinite series discussed in Section 2.2

²Recall that we abbreviate this as GH and that this density is found by physicist's Hermite polynomials; polynomials found by differentiating e^{-x^2}

- $C_{GC}^n(K)$ - the Gram-Charlier option price using only n -terms from its infinite summation.
- $C_H(K)$ - option price under Heston's model.

We will use an adjusted relative error having a denominator greater than 0.01 to prevent numerical errors, this is justified by the fact that options in the United States are never priced below 1 US cent. This relative error is intended to describe how well a given risk-neutral density fits to the aforementioned Heston option price:

$$R(C_{NDF}^m, C_H) = \frac{1}{N} \sum_{j=1}^N \frac{|C_{NDF}^m(K_j) - C_H(K_j)|}{\max(C_H(K_j), 0.01)}.$$

Similarly, to see how well the GC -model will fit to the Heston model we use the formula:

$$R(C_{GC}^n, C_H) = \frac{1}{N} \sum_{j=1}^N \frac{|C_{GC}^n(K_j) - C_H(K_j)|}{\max(C_H(K_j), 0.01)}.$$

To get a numerical Heston option price we use the parameters from Section 4.1.

We performed a 3-fold cross validation on our NDF model; leaving out 1/3rd of our Heston option prices when finding the NDF risk-neutral density coefficients $(a_i)_{i=0}^m$ which were then used to find μ as well as σ . Thereafter we computed the in-sample and out-of-sample relative errors for $m \in \{4, 6, \dots, 14\}$. This was not necessary in the GC -case since, aside from the coefficients itemized above we don't do any fitting; the scope of the GC -expansion in this section is to see how well it *approximates* Heston's model, not how well it fits. Since the out-of-sample and in-sample errors differed only to the 5th decimal place we will only consider the test error. The test errors for the two models is displayed below.

NDF Test Error	GC Test Error	m	n
0.495763409124678	0.108409844389963	4	3
0.0728483878146307	0.0298546397931812	6	4
0.0780088696974434	0.0350818888644778	8	5
0.0388685528271606	0.0810631455081564	10	6
0.151286360809425	0.0591659254461921	12	7
0.157431617652327	0.0762545916270493	14	8

Naturally, for the most part the Gram-Charlier model outperformed the NDF model as it is a cut-off infinite series version of the Heston model. The errors in this model can primarily be attributed to the cutting off and the numerical errors in computing the cumulants from the moment generating function of the Heston model, $M_t(u)$. Moreover, we notice that when $m = 10$ the NDF model gets the best test result. In spite of the high test errors Figure 4.1 shows that our GC and NDF model still do a visually satisfying job of approximating the Heston option price.

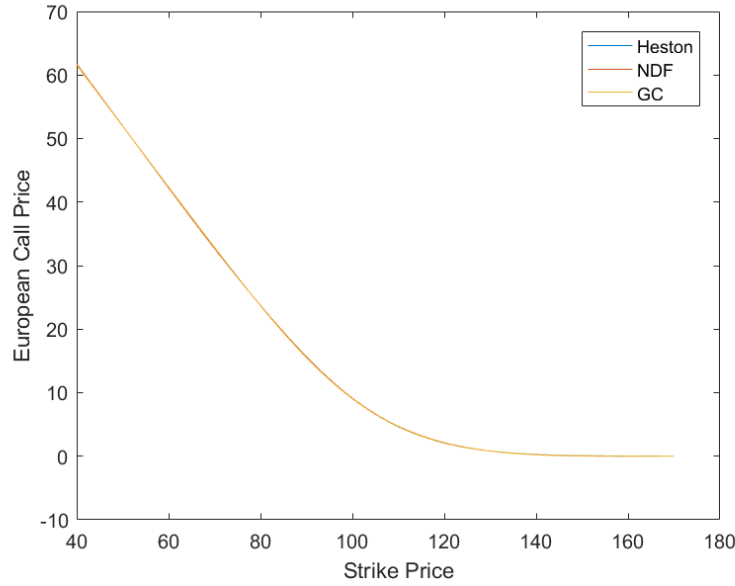


Figure 4.1: Option Prices for Various models, $m=14$, $n=8$

It is worth noting, however, that our model outcome differed from Cheng's. This could, perhaps be due to the fact that we only used the online Mathematica platform to compute our cumulants, which performed worse than our MATLAB computations (explaining why we used MATLAB) whereas Yin presumably used a more sophisticated version of Mathematica. The outcome is displayed in Tables 4.1 and 4.2 below.

Strike	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
50	51.96	51.961	51.961	51.961	51.961	51.961
80	23.714	23.712	23.707	23.693	23.698	23.695
90	15.563	15.528	15.521	15.549	15.564	15.523
100	9.0932	9.0297	9.0286	9.1051	9.1089	9.0655
110	4.6667	4.6093	4.6173	4.679	4.6619	4.6631
120	2.0889	2.0619	2.0731	2.0799	2.0593	2.093
150	0.069433	0.08431	0.084527	0.066061	0.070613	0.071263

Table 4.1: Our Gram-Charlier approximations for the Heston model using the first n cumulants

Strike	$n = 3$	$n = 4$	$n = 5$
50	51.96	51.961	51.961
80	23.714	23.722	23.72
90	15.563	15.551	15.547
100	9.0932	9.062	9.0612
110	4.6667	4.6424	4.6469
120	2.0889	2.0883	2.0939
150	0.0694	0.0893	0.0871

Table 4.2: Cheng's Gram-Charlier approximations for the Heston model using the first n cumulants

Chapter 5

Option Data

A natural next step is seeing how well these models fit to data¹. We select data of call and put options written on Amazon stock on 2/11/2018 expiring at 21/6/2019. The following parameters are used:

- $S_t = 1665.53$ - Amazon asset price at birth of option
- $\tau = 231/365$ - Lifespan of option²
- $r = 0.027$ - The risk-neutral rate³.

For Necula, Drimus, and Farkas' formula it remains to compute the Gauss-Hermite polynomial coefficients for the risk-neutral density through the active-set method for quadratic programming discussed in Section 3.1 as well as the parameters μ and σ through trust region reflective nonlinear least squares from Section 3.2. Additionally, to draw a fair comparison between the two approaches considered we need to fit the parameters of our GC model: $\nu_0, \rho, \kappa, \theta$, and σ to the data.

5.1 Fitting the GC Parameters to the Data

Recall that the GC-model of which we consider is asymptotically equivalent to the Heston model. However, the parameters in our GC option pricing

¹We use data from <https://www.historicaloptiondata.com/content/historical-options-data-file-structures>.

²We are considering a full year as opposed to a trading year in this case.

³We computed this using US 1-year treasury rates at this time https://ycharts.com/indicators/1_year_treasury_rate.

method have a much less explicit relationship with the final option price in comparison to the Heston model. To make things easier we will fit the Heston model to our data and thereafter plug in these coefficients to our GC option pricing formula. As can be seen from the Heston option pricing formula (2.4) we cannot entirely rely on linear nor quadratic programming. Hence we will use the nonlinear least squares method from Section 3.2 to fit this model to our data. Using the shorthand notation for our 5 parameters $\vec{p} \stackrel{def}{=} (\nu_0, \kappa, \sigma, \theta, \rho)$ we let $C_H(K_j; \vec{p})$ denote the price of the Heston call for some strike K_j given the parameters. The objective function for fitting the Heston model to our n data points is:

$$\min_{\vec{p}} \sum_{j=0}^n (C_H(K_j; \vec{p}) - C(K_j))^2.$$

This problem can now be handled by the trust region-reflective least squares algorithm described in Section 3.2.

Filling in gaps of Data Since the data, in itself, does not exhibit a smooth curve we use the cubic splines interpolation method to fill in some of its gaps. We have n data points $(K_j, C(K_j))_{j=1}^n$ and we want to evaluate an intermediate call option price, C , for the m points between $K_j < \tilde{K}_{j1} < \dots < \tilde{K}_{jm} < K_{j+1}$. To do this we consider the system

$$p_j(x) = a_j + b_j(x - K_j) + c_j(x - K_j)^2 + d_j(x - K_j)^3, \quad x \in [K_j, K_{j+1}], \quad j \in \{1, \dots, n-1\}.$$

To find the coefficients $(a_j, b_j, c_j, d_j)_{j=1}^{n-1}$ we need to set up a system of $4(n-1)$ equations. Naturally these curves must intersect with our points:

$$\begin{cases} p_j(K_j) = C(K_j), & j \in \{1, \dots, n-1\} \\ p_{n-1}(K_n) = C(K_n) \\ p_j(K_{j+1}) = p_{j+1}(K_{j+1}), & j \in \{1, \dots, n-2\}. \end{cases}$$

To ensure smoothness and since these functions are three times differentiable we want the first and second derivatives of each p_j to match its successor, p_{j+1} , at each point of intersection:

$$\begin{cases} p'_j(K_{j+1}) = p'_{j+1}(K_{j+1}), & j \in \{1, \dots, n-2\} \\ p''_j(K_{j+1}) = p''_{j+1}(K_{j+1}), & j \in \{1, \dots, n-2\} \end{cases}$$

Finally, we want the second derivatives at the endpoints to be 0

$$\begin{cases} p_1''(K_1) = 0 \\ p_{n-1}''(K_n) = 0, \end{cases}$$

giving us a system of $4(n-1)$ equations with $4(n-1)$ unknowns. To make this implementation we use the MATLAB function *spline* and we evaluate our p -functions for $\tilde{K}_{ji} = K_j + 0.05i$, where $i \in \{1, \dots, \frac{\tilde{K}_{j+1,i} - \tilde{K}_{ji}}{0.05} - 0.05\}$; this produces a much more comprehensive curve than our data alone.

Chapter 6

Data Fitting Results

As done in the previous section we use a 3-fold cross validation to evaluate the performance of our models. For each fold we fit all of the Heston parameters to 2/3rds of the spline adapted data for the Black–Scholes, NDF, and GC models since this yielded a significantly better fit for them than the raw data. However, to achieve the smallest possible relative error in the pure Heston (not GC) setting we used the pure data. Subsequently we compute all our models' respective option prices for the K -values in the test set and evaluate the relative testing error described in the previous results section. Additionally we will consider the cases for the GC model for the first 6 terms as the 7th cumulant yielded a "not a number" value in MATLAB. This is due to the fact that the moment generating function $M_t(u)$ introduced in Section 2.2.1 couldn't be evaluated with the parameter values we acquired from the option data after being differentiated 7 times. Moreover we will consider the coefficients $(a_j)_{j=0}^{n-1}$ where $n \in \{2, 6, \dots, 16\}$. Since the training and test errors were roughly the same for all models we will only discuss the test errors. Having used the tools presented in the previous section the following results were produced.

6.1 NDF Errors

The NDF model yielded inconsistent results since *fmincon* failed to find a minimum on many occasions. This paper only concerns itself when the optima was found.

NDF Test Error	n
0.0713333716082011	2
0.0693943514826509	4
0.00909296454191811	6
0.00914694076555854	8
0.00728423120119476	10
0.00644105230738876	12
0.0064934167710657	14
0.00646823680508747	16

As is displayed in the table above NDF's model converged for high n . At first glance it may seem strange that the model didn't fit best for the highest n when considering that the test and training errors were almost the same (meaning we did not overfit), however, *fmincon* doesn't necessarily stop at the minima; it stops when the difference between our vector of coefficients is sufficiently small between two iterations. This explains why models with more coefficients wouldn't dominate for certain training errors. We display the NDF option prices for $n = 16$ along with the spline fitted data and its relative errors in Figures 6.1 and 6.2 respectively. Furthermore, we fit the Black-Scholes and Heston model to the data using *lsqnonlin* to find the σ in the BS and the numerous parameters in the Heston model that best fits their respective models to the data. The test errors were 0.09 and 0.0098 for Black-Scholes and Heston respectively. Due to its high accuracy the Heston model plotted with the data looks exactly like Figure 6.1. Clearly the NDF and Heston model drastically outperformed the Black-Scholes model, with NDF being the noticeably most accurate model.

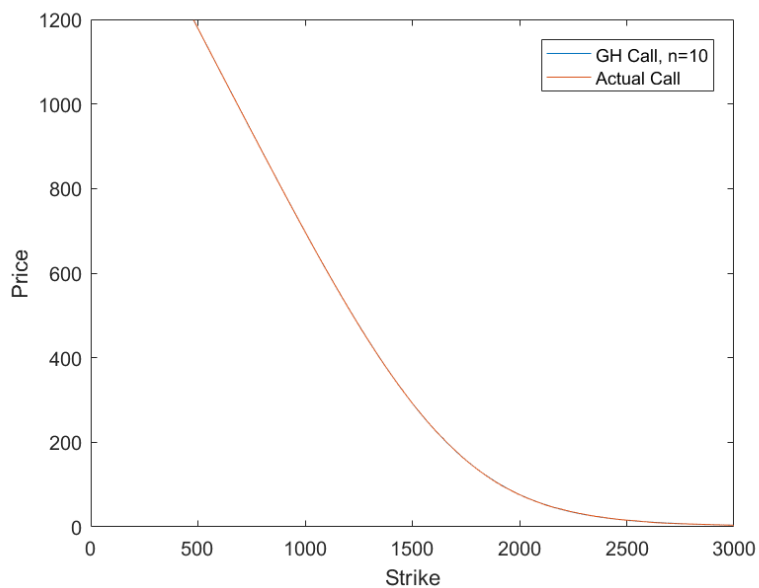


Figure 6.1: NDF option prices plotted with actual option data, $n = 16$

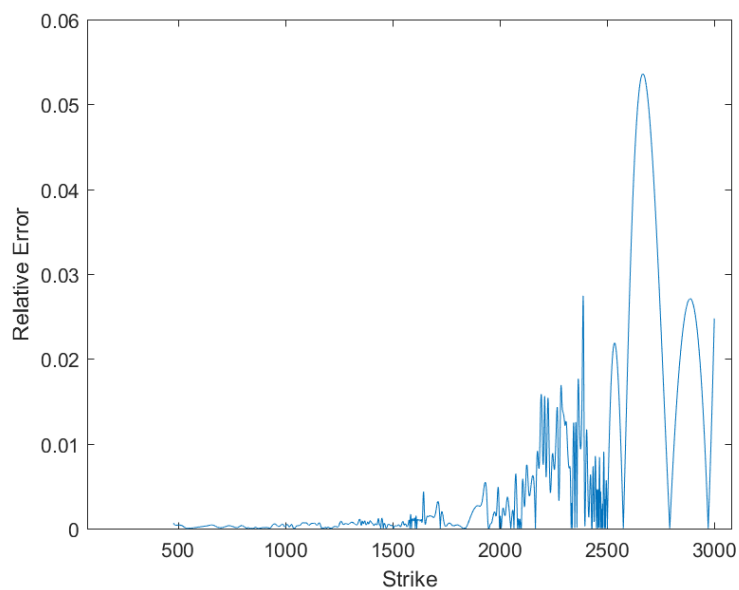


Figure 6.2: NDF relative error, $n = 16$

It is worth noting that Figure 6.2 shows that NDF's model has a much worse performance for deep out-of-the-money options than for other strike prices. This is something that the Black–Scholes model is notorious for. One may think that the reason for this is that in spite of the fact that the Gauss–Hermite distribution which we use for our risk-neutral density isn't normal it is still a polynomial scaling a normal distribution; meaning that it still has a tendency to underestimate the likelihood of rare events. Deep out of the money options only produce a positive return when the asset price soars far beyond its current value which is indeed unlikely, albeit not as unlikely as the normal distribution claims it to be. Furthermore, the Black–Scholes option prices plotted against the spline-fitted option prices and the relative error thereof are displayed in Figures 6.3 and 6.4 respectively. The plots illustrate that the Black–Scholes model struggles especially with out-of-the-money options to an even greater extent than the NDF model. However, upon computing the non-absolute relative error we discover that the Black–Scholes, NDF, and Heston models all overestimated the price of out-of-the-money options. Only our Gram–Charlier expansion drastically underestimated out-of-the-money options because it underestimates the Heston model as well.

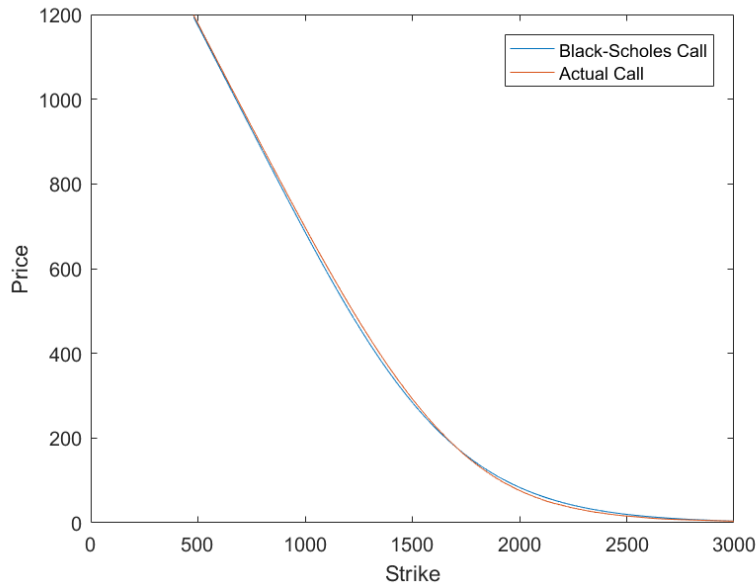


Figure 6.3: Black–Scholes option prices plotted with actual option data.

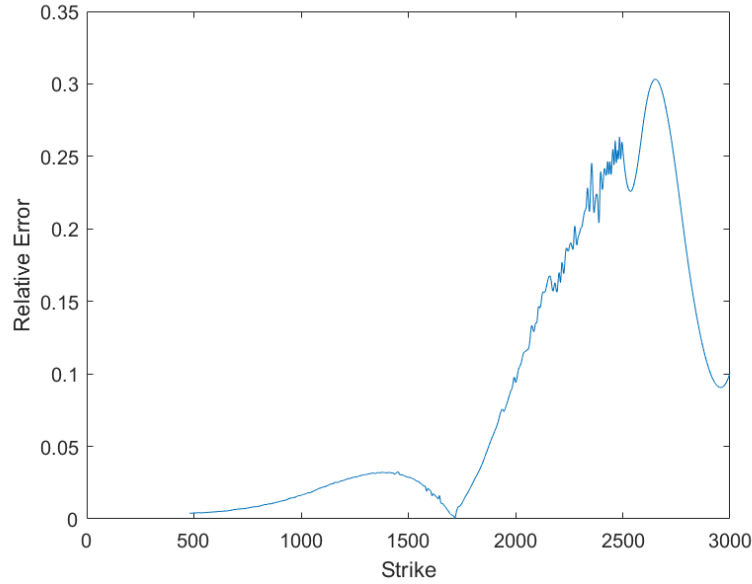


Figure 6.4: Black-Scholes model relative error

6.2 GC errors

As previously mentioned we will cut off our infinite sum from equation (2.8) at points 3, 4, 5 and 6 due to the not-a-number result occurring for the 7th cumulant.

GC Test Errors	n
0.2198345379137871	3
0.259549772655078	4
0.451540429338985	5
0.682847026715799	6

The table above illustrates that the best model in this case was when $n = 3$ as opposed to $n = 5$ in the model section. Recall the arguments made by Corrado and Su [11] supporting the claim that the 3rd moment i.e. the skewness yielded the greatest adjustment to their model and that higher order moments may cause numerical errors. In this case it may mean that

the parameters we typically acquired when tailored to our data produced poor cumulant computations after the third cumulant. The relative errors for this model are an order of magnitude higher than that of NDF's model. Figures 6.6 and 6.7 display how well the model fit along with its relative error respectively. Again, the model performs especially poorly for deep out of the money arguments and takes on a shape similar to that of the Black–Scholes relative error curve, Figure 6.4. It is worth noting that the relative error figures for the Black–Scholes, Gram–Charlier, and NDF models exhibit a smooth curve for higher strikes. This is because we have fewer data points for strikes above 2500, one for every 100, meaning that the error measures how well we fit to the corresponding cubic splines. Figure 6.5 does not exhibit this behaviour since, as mentioned in the beginning of this section, the Heston model fit best to the raw data. Figure 6.5 shows that the Heston model itself also struggled with the deep out of the money options although it still outshined the GC model significantly. It may seem strange that the GC-expansion model was outperformed even by the Black–Scholes model however, this can be explained by the fact that the GC-model relied on the nonlinear least square parameters derived from the Heston model. As we saw in the model fitting results section the GC-expansion model doesn't serve as a particularly good approximation of the Heston model. This means that the Heston model parameters do not necessarily do a good job in fitting the GC-expansion option prices into the data.

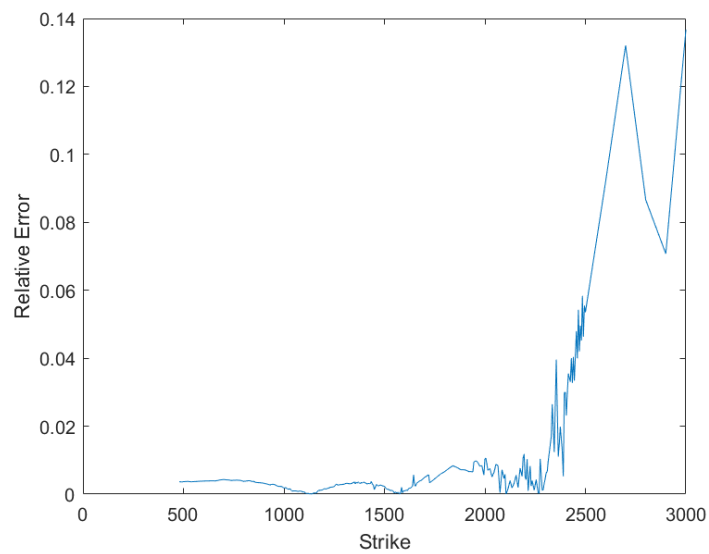


Figure 6.5: Heston model relative error

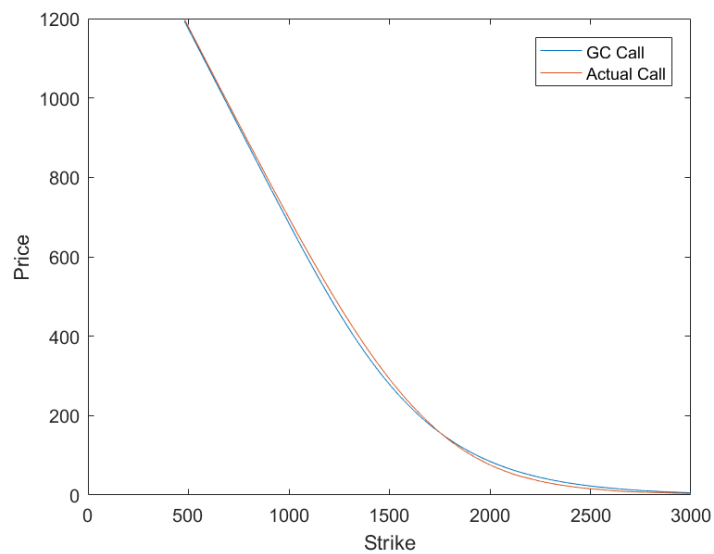


Figure 6.6: GC option prices plotted with actual option data, $n = 3$

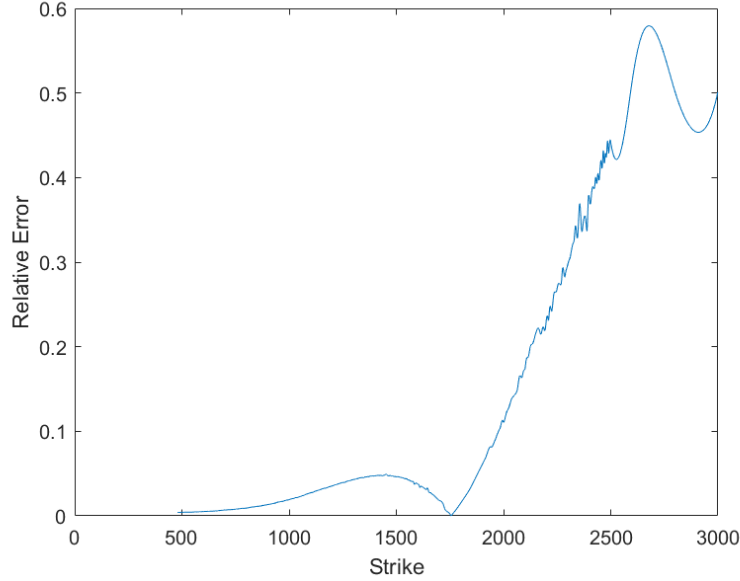


Figure 6.7: GC relative errors, $n = 3$

6.3 Model Complexity

The tables below illustrate the total duration it took to find the parameters for each model and then price the options under them for all the 3 folds. The time to price options under the Gram-Charlier Heston expansion setting consists of the duration of fitting the parameters to the data through *lsqnonlin*, computing the cumulants, and thereafter putting everything together to price the options. The biggest time consumer in the end for this was computing the cumulants, which can be seen in the subsequent table; explaining why the original Heston model was far less time consuming.

GC Time (seconds)	n
11.5413138000000	3
17.6311419900000	4
39.4671034100000	5
112.610722900000	6

Cumulant time (seconds)	n
0.1722814000000000	1
0.4889676000000000	2
1.6621027000000000	3
6.0678931000000000	4
21.6428043000000002	5
73.2235639000000002	6

NDF time (seconds)	n
0.3259234666666667	2
0.9274233	4
0.4394605	6
0.8122020666666667	8
0.570558	10
0.7020970666666667	12
0.9264461	14
1.2315235666666667	16

BS time:0.6302700000000000

Heston time: 12.738389099999999

The NDF model proved to be much more efficient than the other models. Interestingly, in spite of the fact that the Black–Scholes model only had one parameter it was still slower than NDF. This shows how crucial it is for *lsqnonlin* to start at a reasonable point since, in the NDF setting, we only use *lsqnonlin* to acquire the log-mean and standard deviations: μ and σ after having found the a -coefficients of the Gauss–Hermite risk-neutral density. This means that for Necula, Drimus, and Farkas’ model we already begin at an optimum of sorts prior to using non-linear least squares. Additionally, it may seem strange that the time is not monotonically increasing as we increase the number of parameters in the NDF-setting, although this is due to the fact that *fmincon* is fortunate enough to converge quicker for certain starting points within a certain parameter space. These results naturally lead us to conclude that the dominant model is the NDF model for its time efficiency and excellent accuracy.

Chapter 7

Conclusion

7.1 Conclusion

This paper began with a brief overview of Hermite polynomials and how they can be used to derive probability density functions. Thereafter, the fundamental theorem of finance applied to the Heston option pricing setting was discussed with the intention of showing the reader why it is necessary to properly price financial derivatives: to prevent arbitrage opportunities. We illustrated how Hermite polynomials can be used to approximate option prices under the Heston model and to derive a risk-neutral density; a Gauss-Hermite distribution which NDF showed led to a generalization of the Black-Scholes option pricing formula. An explicit formula for deriving the coefficients behind the Gauss-Hermite density from NDF's paper was addressed but then disregarded due to its poor performance. Subsequently we showed how the Gram-Charlier infinite series representation of the Heston model and the NDF model were derived, and proposed methods in quadratic programming and non-linear least squares to fit the parameters of these models to data. We began by fitting the Necula, Drimus, and Farkas option pricing formula and the aforementioned Gram-Charlier expansion to the Heston model which led us to discover that NDF's Gauss-Hermite based model performed nearly as well as the GC-expanded Heston model in fitting something the latter of the two is intended to be an approximation for. The power of NDF's model was further advocated in the following section when it was fit to Amazon option data. The Gram-Charlier expansion, on the other hand, performed relatively poorly due to its second-hand usage of parameters derived for the

sake of fitting the Heston model to the data rendering the GC model to be outperformed by even the single-parameter Black–Scholes model. The NDF model with more than 10 coefficients in the polynomial behind its density was noticeably better than the model with the second best performance; Heston’s model which in turn outperformed the remaining two models having nearly 1/10th of their relative errors. All these models performed at their worse for out-of-the-money options. Finally, the NDF model was interestingly as fast to find its optimal parameter set as the single parameter Black–Scholes model possibly due to the NDF option pricing formula’s corresponding nonlinear least squares problem being more well-behaved caused by the already fit α -coefficients. This is an important detail as the trust region reflective nonlinear least squares method is extraordinarily time consuming.

7.2 Further Studies

Although this paper briefly discussed several other asset pricing models, there is an ocean of viable models which were not explored. An interesting project would be to derive the Gram-Charlier expansions of these alternatives. Furthermore, we only explored European call option prices as these are the simplest and most famous financial derivatives. Of course it is worthwhile to view which model fits a certain type of financial derivative written on a certain type of asset the best or, more importantly, does the best job at predicting its value when it is exercised. Moreover, Necula, Drimus, and Farkas’ model only concerned itself with the mean and standard deviation of log-returns; values corresponding to the 1st and 2nd moments of the asset’s log-price. An interesting experiment would be to consider how to incorporate further moments and to what extent these moments improve or diminish their model’s ability to fit to data in a fashion similar to that explored in our GC option pricing experiments.

Bibliography

- [1] S. Heston, “A closed-form solution for options with stochastic volatility with applications to bond and currency options,” *The Review of Financial Studies*, vol. 6, no. 2, pp. 327–343, 1993.
- [2] Y. H. Cheng, “Pricing derivatives by gram-charlier expansions,” 2013.
- [3] C. Necula, G. Drimus, and W. Farkas, “A general closed form option pricing formula,” *Review of Derivatives Research*, vol. 22, p. 1–40, 2017.
- [4] F. Black and M. Scholes, “The pricing of option and corporate liabilities,” *The Journal of Political Economy*, vol. 81, no. 3, p. 637–654, 1973.
- [5] A. Hald, “The early history of the cumulants and the gram-charlier series,” *The Review of Financial Studies*, vol. 68, no. 2, pp. 137–153, 2000.
- [6] R. Jarrow and A. Rudd, “Approximate option valuation for arbitrary stochastic processes*,” *Journal of Financial Economics*, vol. 10, pp. 347–369, 1982.
- [7] J. Cox, “Notes on option pricing i: Constant elasticity of diffusions,” 1975.
- [8] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” *Econometrica*, vol. 50, pp. 987–1008, 1982.
- [9] T. Bollerslev, “Generalized autoregressive conditional heteroscedasticity,” *Econometrica*, vol. 31, no. 3, pp. 307–327, 1986.
- [10] P. Hagan, D. Kumar, A. Lesniewski, and D. Woodward, “Managing smile risk,” *Wilmott Magazine*, vol. 1, pp. 84–108, 2002.

- [11] C. Corrado and T. Su, “S&p 500 index option tests of jarow and rudd’s approximate option valuation formula,” *The Journal of Futures Markets*, vol. 16, no. 6, pp. 611–629, 1996.
- [12] H. Cramer, “Mathematical methods of statistics,” 1957.
- [13] J. Gatheral, *The Volatility Surface: A Practitioner’s Guide*. Wiley Finance Series, 2006.
- [14] S. Del Baño Rollin, A. Ferreira-Castilla, and F. Utzet, “A new look at the heston characteristic function,” *Mathematics Subject Classification*, 2009.
- [15] W. Myller-Lebedeff, “Die theorie der integralgleichungen in anwendung auf einige reihenentwicklungen,” *Mathematische Annalen*, vol. 64, pp. 388–416, 1907.
- [16] G. Bakshi, N. Kapadia, and D. Madan, “Stock return characteristics, skew laws, and the differential pricing of individual equity options,” *The Review of Financial Studies*, vol. 16, pp. 101–143, 2003.
- [17] J. Nocedal and S. Wright, *Numerical Optimization*. Springer, 2006.