The Calibration of Option Pricing Models

Hilmar Haukur Gudmundsson

Supervisor: prof. dr. David Vyncke
Dissertation submitted in fulfillment of the requirements for the degree of
Doctor in Science: Mathematics

Department of Applied Mathematics,
Computer Science and Statistics
Faculty of Sciences
Ghent University
Academic year 2019-2020

Members of the examination committee

Prof. dr. Kris Boudt
Department of Economics
Ghent University, Belgium

Prof. dr. Marnix Van Daele (chair)
Department of Applied Mathematics,
Computer Science and Statistics
Ghent University, Belgium

Dr. Kathrin Glau
School of Mathematical Sciences
Queen Mary University of London, UK

Prof. dr. Karel in 't Hout
Department of Mathematics
University of Antwerp, Belgium

Prof. dr. Michèle Vanmaele
Department of Applied Mathematics,
Computer Science and Statistics
Ghent University, Belgium

Prof. dr. David Vyncke (promotor)
Department of Applied Mathematics,
Computer Science and Statistics
Ghent University, Belgium

Acknowledgements

First, I would like to express my sincere gratitude to my advisor Prof. David Vyncke for his support. He gave me the flexibility I needed to experiment with new ideas, and his expert opinion when I needed help. His patience with my long rants about my ideas in his office was greatly appreciated, his commonsense feedback kept me sane and his knowledge of mathematical finance kept me on my toes. I could not have asked for a better advisor throughout these years of research.

I would also like to thank members of my thesis committee: Prof. Michèle Vanmaele, Prof. Karel in 't Hout, Prof. Kris Boudt and Dr. Kathrin Glau for their insightful comments and probing questions which gave me a chance to see aspects of my research in a new light.

My sincere thanks also go to Prof. Einar Steingrímsson and Prof. Sergey Kitaev for completely transforming the way I think about mathematics during my first two years in graduate school. Their mentoring largely shaped how I approach mathematical research to this day. I am also grateful to Dr. Jón Friðrik Sigurðsson for skillfully putting me on the right track during challenging times in my life.

I also thank my friends and colleagues: Johan, Vahe, Bashir, Xianming, Paula, Camila, Mushthofa, Christophe, Oliver, Bea, Diego, Koen, Fatemeh, Catherine, Hans, Mina and others for sharing the laughs, beers, troubles and triumphs with me.

My mom and dad have my deepest gratitude for believing in me and supporting me throughout all these years of school. Knowing that they would always be there for me gave me the strength to keep going.

And finally, I want to thank my beloved Elena for keeping my feet on the ground with her great sense of humor and broad intellect, for inspiring me to become a better man, and for reminding me that there is more to life than math and programming.

Hilmar Gudmundsson
Ghent, December 16, 2019

Contents

List of Figures

List of Tables

List of Abbreviations

BAW Barone-Adesi-Whaley. 80, 81

BFGS Broyden-Fletcher-Goldfarb-Shanno. 88, 102, 105

CARA constant absolute risk aversion. 22

CIR Cox-Ingersoll-Ross. 15

COS Fourier-cosine series expansion. 39, 40, 51, 52, 53, 54, 64

CPT Cumulative Prospect Theory. 86, 93, 94, 96, 97

CRRA constant relative risk aversion. 22

D32 double 3/2 model. 61, 62, 63

EM Euler-Maruyama. 23, 24

FFT fast Fourier transform. 34, 38, 39, 40, 51, 52, 53, 54, 64

FFTAP First Fundamental Theorem of Asset Pricing. 17

GARCH generalized autoregressive conditional heteroskedasticity. 45

GBM geometric Brownian motion. 14, 95, 108, 109, 111

GMM generalized method of moments. 45

GMSC gradient-maturity-strike caching. 42, 43, 51, 52, 53, 54, 61, 62, 63

GWMC generalized weighted Monte Carlo. 99, 100, 101, 105, 107, 108, 109, 111

HARA hyperbolic absolute risk aversion. 22

MAPE mean average price error. 101, 106, 107

MCMC Markov chain Monte Carlo. 4, 6, 31, 32, 46, 47, 48, 50, 57, 64, 66

MGMSC mixed gradient-maturity-strike caching. 63

MH32 mixed Heston-3/2 model. 62, 63

MRPE mean relative price error. 101, 106, 107, 111

OTC over-the-counter. 1

OTM out-of-the-money. 79, 93, 96, 107, 111

OWMC original weighted Monte Carlo. 99, 104, 105, 107, 108, 109, 111

PDE partial differential equation. 14, 23, 67, 80

RMSIV root mean squared implied volatility. 81

RMSP root mean squared price. 81

SC strike caching. 41, 51, 52, 53, 54, 62, 63, 64

SFTAP Second Fundamental Theorem of Asset Pricing. 17

WMC weighted Monte Carlo. 85, 86, 87, 91, 112

Samenvatting

Het onderzoek in dit proefschrift richt zich op verschillende aspecten van de kalibratie van optieprijsmodellen. In het eerste hoofdstuk wordt het kalibratieprobleem gesitueerd in de wetenschappelijke literatuur. Het tweede hoofdstuk geeft vervolgens een overzicht van de fundamenten uit de stochastische analyse, de financiële economie en de computationele financiële wiskunde waarop de resultaten uit de volgende hoofdstukken steunen.

In het derde hoofdstuk, getiteld *On the Calibration of the 3/2 Model*, wordt een kalibratiemethode voor het 3/2 stochastische volatiliteitsmodel voorgesteld. Daarbij wordt gebruikgemaakt van de analytische gradiënt van de karakteristieke functie van het model, samen met een caching algoritme dat het kalibratieproces van het model aanzienlijk versnelt. Daarnaast wordt een regularisatie van de kalibratiedoelfunctie voorgesteld waarbij het regularisatiepunt en de dempingsmatrix worden verkregen uit een MCMC-schatting van de tijdreeks van de geïmpliceerde volatiliteit.

In het vierde hoofdstuk, getiteld *Non-Affine Stochastic Volatility With Seasonal Trends*, wordt een algemener 3/2 model voorgesteld dat de seizoensgebonden trends in de volatiliteit van futures op granen beschrijft. De belangrijkste bijdrage – naast de nieuwe modeldynamiek op zich – is de afleiding van de karakteristieke functie van het model in analytische vorm, wat de kalibratie van het model aanzienlijk vergemakkelijkt. Daarnaast wordt een kalibratiemethode met een hybride globaal-lokale optimalisatieprocedure voor het minimaliseren van de kleinste-kwadratendoelfunctie uitgewerkt. Pure gradiëntgebaseerde methoden zijn immers problematisch door het gebrek aan strikte convexiteit in het optimalisatieprobleem. De kalibratieprocedure werd vervolgens op een parallel computerplatform (CUDA) geïmplementeerd om ze te versnellen. De numerieke tests tonen aan dat dit nieuwe model in staat is om de prijzen van opties op futures op maïs en tarwe met een aanzienlijk grotere nauwkeurigheid te reproduceren dan het originele 3/2 model.

In het vijfde hoofdstuk, getiteld *A Generalized Weighted Monte Carlo Calibration Method*, wordt een veralgemeende versie van de gewogen Monte Carlo kalibratiemethode voorgesteld. De oorspronkelijke methode berekent gewichten voor de Monte Carlo paden van het te kalibreren arbitragemodel zodat de gewogen paden exact de marktprijzen van de benchmarkopties reproduceren en bovendien zodanig dat de gewichten een minimale relatieve entropie vertonen ten opzichte van de oorspronkelijke (uniforme) gewichten. De veralgemening van deze methode omvat een vervorming van de staartkansen, evenals een opsplitsing van de gegenereerde paden op basis van de looptijd van de opties zodat elk segment van het pad een eigen gewicht krijgt. Een uitgebreide numerieke studie toont aan dat deze nieuwe methode tot aanzienlijke verbeteringen in de out-of-sample fit leidt voor optieprijzen geschreven op de S&P 500 voor twee veelgebruikte optieprijsmodellen. Het hoofdstuk bevat ook een gedetailleerde bespreking van de wiskundig equivalente interpretatie van de gewogen Monte Carlo methode als een optimalisatieprobleem voor een portefeuille van een belegger met een verwacht nut.

Summary

The research in this dissertation focuses on different aspects of the calibration of option pricing models. In the first chapter the calibration problem is situated in the scientific literature. The second chapter gives an overview of the concepts from stochastic analysis, financial economics and computational finance, serving as foundations for the results in later chapters.

In chapter 3, titled *On the Calibration of the 3/2 Model*, a calibration method is proposed for the 3/2 stochastic volatility model which uses the analytic gradient I derive for the characteristic function of the model combined with a caching algorithm that greatly speeds up the calibration process of the model. In addition, a regularization of the calibration objective function is proposed where the regularization point and the damping matrix are obtained from MCMC estimation of the time series of past implied volatility.

In chapter 4, titled *Non-Affine Stochastic Volatility With Seasonal Trends*, a more general 3/2 model is proposed which captures the seasonal trends in the volatility of commodity futures on grains. The main contribution besides the new model dynamics is the derivation of the characteristic function of the model in closed form, which greatly facilitates the calibration of the model. In addition, a calibration procedure for the model is presented that uses a hybrid global-local optimization routine for finding the minimum of the least squares objective function, which tackles the lack of strict convexity present in the problem that can make purely gradient-based methods problematic. Furthermore, I implemented this calibration procedure on a parallel computing platform (CUDA) to speed it up. The numerical tests of this new model reveal that it is capable of reproducing the prices of options on futures on corn and wheat with significantly greater degree of accuracy than the original 3/2 model.

In chapter 5, titled *A Generalized Weighted Monte Carlo Calibration Method*, a generalized version of the weighted Monte Carlo calibration method is proposed. The original method computes a set of weights for the Monte Carlo sample paths for the no-arbitrage model to be calibrated such that

the reweighted paths reproduce exactly the market prices of the calibration benchmark options and furthermore such that the weights exhibit minimum relative entropy with respect to the prior (uniform) weights. The generalization of this method proposed here incorporates a distortion of the probability of tail events as well as a maturity-based partition of the sample paths with each segment of each path getting its own weight. The results of extensive numerical tests are presented which show that this new method leads to significant improvements in the out-of-sample fit in terms of the prices of options written on the S&P 500 for the two commonly used option pricing models that were calibrated. The chapter also includes a detailed discussion on the mathematically equivalent interpretation of the weighted Monte Carlo method as a portfolio choice problem for an investor with expected utility.

Chapter 1

Introduction

Financial derivatives constitute a broad class of contingent claims that are used for reallocating capital across uncertain future outcomes. The main theoretical justification behind financial derivatives is that they allow market participants to hedge against financial risk, essentially making financial derivatives a class of insurance instruments. Modern financial economics carries the fundamental assumption that participants in financial markets are risk averse, in the sense that risky assets trade at a discount relative to their expected value in the market. Since risk aversion among capital holders is commonly understood to limit the access firms and individuals have to investment capital, instruments that allow capital holders to share risk more effectively are implicated as a requirement for increasing the efficiency of capital allocation in the economy.

Although the derivatives industry has experienced several setbacks in recent history, the market demand for derivatives continues to be significant. At the time this is written, the Bank for International Settlements estimates the combined market value of outstanding over-the-counter (OTC) derivative contracts to be roughly 9.7 trillion dollars (BIS [2019]).

While derivative contracts have been traded as far back as several thousand years ago (Crawford and Sen [1996]), the modern mathematical framework of option pricing is generally traced back to the work of Fischer Black and Myron Scholes who in 1973 published a simple option pricing model which still remains widely used today (Black and Scholes [1973]).

The research literature on derivative pricing now contains a vast and diverse set of market models which improve upon the original Black-Scholes model in a variety of ways. In fact, modern no-arbitrage models have been shown to be able to capture the empirical features of the price processes of stocks and indices to such a high degree that choosing between them now

requires considerations beyond their basic empirical fit to the calibration instruments (Schoutens et al. [2005]).

Meanwhile, the fundamental question of how to efficiently calibrate these models still remains largely open. One reason is that market models that are flexible enough to capture the dynamics of financial markets tend to be too mathematically intractable for standard statistical inference methods to be applicable. Furthermore, if a model is misspecified, or if the data is noisy, the calibration process carries with it a subjective judgment on how the loss function should be formulated to produce the most plausible parameter estimates.

The calibration of a market model refers here to the procedure of computing admissible values for the model parameters so that the derivative prices implied by the model coincide with the corresponding benchmark derivative prices observed in the market. The standard approach to calibrating an option pricing model given a set of observed cross-sectional derivative prices in the market is through the minimization of the model prediction error. It amounts to obtaining

$$\boldsymbol{\theta}^* = \text{argmin}_{\boldsymbol{\theta}} \sum_i w_i ||h(\pi_i^{\text{market}}, \pi_i^{\text{model}}(\boldsymbol{\theta}))||, \tag{1.1}$$

where π_i^{market} is the observed market value for contract i, $\pi_i^{\text{model}}(\boldsymbol{\theta})$ is the price predicted for contract i by the model given the parameter vector $\boldsymbol{\theta}$, w_i is the weight that determines how much the discrepancy between the market price and model price for contract k influences the calibration objective function, h is the form of the prediction error, and $||\cdot||$ is the norm under which the magnitude of the error is calculated. The summand $||h(\cdot)||$ is generally referred to as the loss function.

No decisive economic intuition exists so far on which loss function to use to quantify the discrepancy between the model prices and the observed prices. Some amount of research has been dedicated to answering the question on the basis of general consistency. A handful of authors have attempted to deal systematically with the choice of different loss functions when computing (1.1). Christoffersen and Jacobs [2004] argue that the loss function for the in-sample estimation and the function for the out-of-sample evaluation of a model should be the same to obtain a consistent evaluation of the model. Bams et al. [2009] propose a method for evaluating the relative performance of different loss functions and argue that the interaction between different models and loss functions should be taken into consideration when analyzing the pricing performance. Friedman et al. [2014] propose utility-based loss functions to make the prediction error more economically meaningful.

The problem given by (1.1) falls under the category of inverse problems. Within that framework, it is in the general case an ill-conditioned problem, in the sense that small changes to the option price data can lead to large changes in $\boldsymbol{\theta}^*$ (Cont [2010]). This is especially pertinent since option price data tends to be noisy. To begin with, option prices are usually not available in the market except as a bid-ask interval. A typical resolution to this is simply taking the average of the bid and ask price, however there is strictly speaking no guarantee that this is the fair market price of the option. Secondly, the liquidity of option contracts tends to fall off rapidly as either the strike price moves away from the price of the underlying, or as the option maturity moves beyond the span of a few months. An econometrically precise definition of liquidity is still lacking in the literature[1], but it is generally assumed that a contract needs to be liquid in order for its price to be an accurate measure of the pricing function of the market. Other sources of noise include fixed "easy number" increments of the option prices with respect to the strike price, and the asynchronous nature of the price observations.

Aside from the calibration problem being ill-conditioned, the associated objective function often appears to lack strict convexity (Hirsa [2016]; Gilli and Schumann [2011]; Guillaume and Schoutens [2010]; Cont and Tankov [2004]; Gudmundsson and Vyncke [2019a]). This complicates using gradient-based methods for solving for (1.1) since the optimization routine can get stuck in a local minimum.

With these issues in mind, several authors have proposed more robust calibration methods for retrieving (1.1). A method commonly used for stabilizing inverse problems is adding a convexifying penalty term to the calibration objective function. In finance, this approach has so far mostly been studied in the context of local volatility models (see e.g. Crépey [2010]; Egger and Engl [2005]; Cezaro et al. [2010]). Apart from utilization in local volatility modeling, Cont and Tankov [2004] propose a regularized calibration objective for exponential Levy models that has a penalty term derived from relative entropy.

Another approach to offset the noise in purely cross-sectional option data is to add historical prices to the calibration procedure (Bates [2003]). The most ambitious version of this is a full joint estimation of the parameters of the model under both the physical and the risk neutral measure. This means incorporating the full time series of both option prices and the price of the underlying into the estimation procedure. This tends to be an enormously

1. Traded volume (as well as open interest) is often taken as a proxy to the liquidity of an option but a glance at the options listed on the S&P500 reveals that on a given day contracts with zero trading volume are often sandwiched between option contracts with a large trading volume, implying that this is a somewhat questionable measure of liquidity.

challenging task from a computational standpoint since most option pricing models do not yield a known transition density in closed form.

However, the problem has attracted significant attention in the field of econometrics in recent years. The variety of methods proposed in this context are too numerous and broad to delve into here, but major avenues of research include likelihood function approximation (Aït-Sahalia and Kimmel [2007]), generalized method of moments (Pan [2002]), method of moments with semi non-parametric auxiliary models (Chernov and Ghysels [2002]), and Markov chain Monte Carlo (MCMC) (Eraker [2004], Jones [2003]).

Sidestepping the computational challenges of full joint estimation, some authors have proposed estimating a subset of the model parameters first using historical data, thereby reducing the parameter space that the optimization routine needs to consider for solving the calibration problem given by (1.1). Guillaume and Schoutens [2010] propose using the time series of historical implied volatility to get an estimate for the long run mean and instantaneous volatility of the Heston model, which confines the calibration algorithm to a proper subset of the model parameters that turn out to be more easily identifiable. Arismendi et al. [2016] construct a seasonal modification of the Heston model which they estimate under the physical measure using an MCMC estimator, and the estimates are then plugged into the model which only leaves the parameter corresponding to the market price of volatility risk to be determined in the cross-sectional part of the calibration.

The application of Bayesian estimation to option pricing models for the sake of robustness is also the topic in Gupta and Reisinger [2014] where it is applied to a local volatility option pricing model to obtain a range of parameter estimates for which the model is able to replicate observed option prices such that they fall within a predetermined error band.

A more active approach to the calibration problem was proposed in Avellaneda et al. [2001] where the authors essentially transform the calibration into a model correction procedure referred to as the weighted Monte Carlo method. Friedman et al. [2013] refine the method by various means such as t-distribution innovations for the path simulation and a new kind of entropy measure.

The statistical reliability of the parameter estimates is only one aspect of the calibration process. The other major consideration is the speed and numerical accuracy with which we can obtain them. While we can reasonably expect the pricing of a single derivative to be achieved well under a second on most modern computer systems, the calibration procedure for the pricing model can require many thousands of option price evaluations.

This fact serves as a motivation for choosing models for which we can derive (semi)closed form formulas for plain option pricing.

Assuming the loss function is sufficiently smooth with respect to the model parameters, the most numerically efficient way of computing (1.1) will involve second-order gradient-based optimization. In the event that the calibration objective function is found to be non-convex and regularization is unavailable, such methods can be coupled with heuristic global optimization algorithms to avoid local minima. However, these tend to require a significantly greater number of iterations than purely gradient based ones, so their application generally requires a highly streamlined implementation to be computationally feasible.

Nonlinear optimization has been studied for centuries with a research literature to match. For a general overview see e.g. Nocedal and Wright [2006]. A few contributions have been made that focus solely on the optimization aspect of the calibration problem in finance. Gerlich et al. [2012] propose a feasible point sequential quadratic programming algorithm which they test on the Heston model. Cui et al. [2017] derive the gradient of the characteristic function of the Heston model which they use with a caching algorithm to greatly speed up the calibration of the model. Yang and Lee [2012] compute the gradient of the characteristic function of several Levy models which they use to speed up an implementation of the particle swarm method for solving the calibration problem of those models.

When the model in question yields a probability density for which there is a known characteristic function, the speed aspect of the calibration is also impacted by the numerical methods we use for the Fourier inversion part. Several methods have been proposed for efficient ways of Fourier pricing specifically for the model calibration setting when many options need to be priced simultaneously. Carr and Madan [1999] propose using the fast Fourier transform method for simultaneously generating the prices of the same option for many different strike prices. Chourdakis [2005] implements the fractional fast Fourier Transform for the calibration problem and tests it on both stochastic volatility and variance gamma models. Kilin [2011] proposes an inversion algorithm that caches the intermediate characteristic function evaluations in between pricing different options and tests it on stochastic volatility models. An inversion method based on a Fourier-cosine expansion is proposed in Fang and Oosterlee [2008] where the inversion integral is replaced by its cosine series expansion. And in Gaß et al. [2017] an offline-online procedure is proposed for constructing efficient numerical quadrature rules for computing the Fourier integral.

The research presented in this thesis concerns several different aspects of the calibration procedure of option pricing models, with emphasis on the computational perspective. In chapter 3, which is based on the published paper Gudmundsson and Vyncke [2019a], the focus is on the 3/2 stochastic volatility model which is unique in the sense that it is a non-affine stochastic volatility model with a known characteristic function. We take advantage of this fact to compute the gradient of the model's characteristic function in closed form and propose further a caching caching algorithm which enables us to calibrate the model more than 10 times faster than the fastest comparison method we could find in the literature. In the second half of chapter 3 we derive MCMC estimators for the 3/2 model for the purpose of proposing a type of regularization for the least squares objective function which we termed "risk neutral regularization", which refers to the fact that the estimation is contained entirely within the risk-neutral probability measure. These methods are tested both on simulated as well as real S&P500 options data.

In chapter 4, which is based on the working paper Gudmundsson and Vyncke [2019b], the focus continues to be on the 3/2 volatility specification, however we extend the model by adding to it a cyclical volatility trend which we show empirically to be well suited for capturing known seasonal effects in certain types of futures markets. We derive the characteristic function for this new model in closed form. Furthermore, we test it on real CBOE options data using a calibration procedure which combines global and local search methods for added robustness, as well as parallelization of the computation of the least squares objective function across GPU threads for added speed.

In chapter 5, which is based on the working paper Gudmundsson and Vyncke [2019c], we go beyond the scope of the 3/2 volatility specification and propose a general extension to the weighted Monte Carlo calibration method of Avellaneda et al. [2001]. This generalization introduces a probability distortion that is applied to the prior measure of the weighted Monte Carlo method (which so far has been taken as uniform by default). The probability distortion is inspired by elements of Cumulative Prospect Theory (Kahneman and Tversky [1992]), and we show through extensive numerical testing on S&P500 options data that this modified weighted Monte Carlo method is capable of calibrating the Monte Carlo paths in such a way that they produce a much better fit to out-of-sample options prices than what can be achieved through the original method.

Chapter 2

Theoretical Preliminaries

In this chapter we establish some terminology and review fundamental definitions and well known results from both stochastic analysis, financial economics and computational finance which we can think of as being a foundation on top of which the work presented in later chapters is built. Reference material includes Chesney et al. [2009] for sections 2.1 and 2.2; Back [2010], Downarowicz [2010] and Biagini [2010] for section 2.3; Skiadas [2009] for section 2.4; Glasserman [2013] for section 2.5.1; and Schmelzle [2010] for section 2.5.2. We point to these references for proofs and further elaboration of the results presented in this technical review.

2.1 Measure-Theoretic Probability

We begin with the relevant definitions from measure-theoretic probability theory.

Definition 2.1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A **filtration** $\mathbb{F}$ on $(\Omega, \mathscr{F}, \mathbb{P})$ is an increasing family $(\mathscr{F}_t)_{t\geq 0}$ of sub-σ-algebras on $\mathscr{F}$. More specifically, for each t, $\mathscr{F}_t$ is a σ-algebra included in $\mathscr{F}$ and if $s \leq t$ then $\mathscr{F}_s \subseteq \mathscr{F}_t$. The tuple $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ is referred to as a **filtered probability space**.*

Throughout this thesis, we assume that every filtered probability space has a filtration that satisfies the *usual conditions*.

Definition 2.2. *A filtration $\mathbb{F}$ on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ is said to satisfy the **usual conditions** if*

1. *$\mathscr{F}_0$ (hence every $\mathscr{F}_t$) is complete relative to $\mathscr{F}$ with respect to $\mathbb{P}$ in the sense that $\mathscr{F}_0$ contains every set A that is contained in the power set of Ω such that $\mathbb{P}(A) = 0$.*

2. $\mathbb{F}$ *is right continuous, i.e.,* $\mathscr{F}_t = \mathscr{F}_{t+}$ *for all* $t \geq 0$*, where*

$$\mathscr{F}_{t+} \equiv \bigcap_{s>t} \mathscr{F}_s,$$

Changing the measure with respect to which we compute the expectation of random variables is a common practice in the field of option pricing. This is because we generally want to abstract our models away from containing explicit terms involving the market price of risk. We start with the definition of equivalence between measures.

Definition 2.3. *Let* $\mathbb{P}$ *and* $\mathbb{Q}$ *be two probability measures constructed on the same measurable space* $(\Omega, \mathscr{F})$*. We say that* $\mathbb{Q}$ *is* ***absolutely continuous*** *with respect to* $\mathbb{P}$ *(usually denoted as* $\mathbb{Q} \ll \mathbb{P}$*) if for every* $A \in \mathscr{F}$ *such that* $\mathbb{P}(A) = 0$ *we have that* $\mathbb{Q}(A) = 0$*. Furthermore, we say that* $\mathbb{P}$ *and* $\mathbb{Q}$ *are* ***equivalent*** *if they have the same set of impossible events i.e.,*

$$\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0,$$

for all $A \in \mathscr{F}$*.*

The next theorem, usually referred to as the Radon-Nikodym theorem, is fundamental in measure and integration theory. As we explain shortly, it is particularly relevant to finance as well.

Theorem 2.4. *Let* $\mathbb{P}$ *and* $\mathbb{Q}$ *be two measures on* $(\Omega, \mathscr{F})$ *and* $\mathbb{P}$ *be* σ*-finite. If* $\mathbb{Q} \ll \mathbb{P}$*, i.e., if* $\mathbb{Q}$ *is absolutely continuous with respect to* $\mathbb{P}$*, then there exists a nonnegative Borel function* f *on* Ω *such that for every* $A \in \mathscr{F}$ *it holds that* $\mathbb{Q}(A) = \int_A f \, d\mathbb{P}$ *The function* f *is referred to as the Radon-Nikodym derivative of* $\mathbb{Q}$ *with respect to* $\mathbb{P}$ *and is denoted by* $\frac{d\mathbb{Q}}{d\mathbb{P}}$*.*

Our main object of study is the stochastic process. We can think of it as a time-indexed sequence of random variables. More formally, we have the following definition.

Definition 2.5. *A* ***stochastic process*** $X = (X_t)_{t \geq 0}$ *is a family of mappings from the statespace* Ω *into* $\mathbb{R}$*. Furthermore,* X *is said to be* ***continuous*** *if the map* $t \to X_t(\omega)$ *is continuous.*

Definition 2.6. *A stochastic process* $X = (X_t)_{t \geq 0}$ *on* $(\Omega, \mathscr{F}, \boldsymbol{F}, \mathbb{P})$ *is said to be* ***adapted*** *to the filtration* $\mathbb{F}$ *if, for each* $t \geq 0$ *we have that* X_t *is* $\mathscr{F}_t$*-measurable.*

In discrete time, we can intuitively describe a stochastic process $(X_n)_{n \in \mathbb{N}}$ as being *predictable* if X_{n+1} is measurable with respect to $\mathscr{F}_n$ for each n. In other words, at time n we already know the value of X_{n+1}. This has a

straightforward interpretation in finance where we can think of trading strategies in a discrete, dynamic setting as predictable processes. In continuous time, the definition becomes more involved, where the predictability trait is captured by left-continuity.

Definition 2.7. *Let $\mathscr{P}$ denote the smallest σ-algebra on $[0,\dots,T]\times\Omega$ that contains every adapted left-continuous process on the state-space. A stochastic process that is measurable with respect to $\mathscr{P}$ is said to be* ***predictable****.*

The concept of progressive measurability is important for the definition of stochastic integrals which we give below.

Definition 2.8. *A stochastic process $X=(X_t)_{t\geq 0}$ is said to be* ***progressively measurable*** *with respect to the filtration $\mathbb{F}$ if, for every t, the map $[0,t]\times\Omega\to\mathbb{R}$ defined by $(s,\omega)\to X_s(\omega)$ is measurable with respect to the product σ-algebra $\mathscr{B}[0,t]\times\mathscr{F}_t$ where $\mathscr{B}[0,t]$ is the Borel σ-algebra on $[0,t]$.*

One general property that is embedded in almost every stochastic process used in continuous-time finance is the Markov property. While it is not implied intrinsically by an arbitrage free market, it is a major requirement for tractability.

Definition 2.9. *Let $X=(X_t)_{t\geq 0}$ be an adapted stochastic process, with $\mathscr{F}_t^X=\sigma(X_s,s\leq t)$. The process X is referred to as a* ***Markov process*** *if for any times $t\geq s\geq 0$ and any bounded Borel function f we have that*

$$\mathbb{E}[f(X_t)|\mathscr{F}_s^X]=\mathbb{E}[f(X_t)|X_s],$$

In other words, a Markov process's future state is stochastically independent of its past given its present state.

Finally, the following definitions will be used throughout the remainder of this chapter.

Definition 2.10. *A stochastic process $X=(X_t)_{t\geq 0}$ is said to be* ***integrable*** *if for all $t\geq 0$ we have that $\mathbb{E}[|X_t|]<\infty$. Furthermore, X is said to be* ***square integrable*** *if for all $t\geq 0$ we have that $\mathbb{E}[X_t^2]<\infty$. In particular, we denote with $\mathscr{L}_2$ the set of all square integrable processes.*

2.1.1 Martingales

The concept of martingales is central to mathematical finance since they represent "fair bets" and are therefore closely connected to the assumption that markets are arbitrage free.

Definition 2.11. *If the adapted stochastic process $X = (X_t)_{t\geq 0}$ is integrable, then it is called*

- *A **martingale** if $\mathbb{E}[X_t|\mathscr{F}_s] = X_s$ for all $0 \leq s \leq t$,*
- *A **submartingale** if $\mathbb{E}[X_t|\mathscr{F}_s] \geq X_s$ for all $0 \leq s \leq t$,*
- *A **supermartingale** if $\mathbb{E}[X_t|\mathscr{F}_s] \leq X_s$ for all $0 \leq s \leq t$.*

A broader class of stochastic processes is the class of *local martingales*, for which we need to introduce the concept of stopping times.

Definition 2.12. *A random variable $\tau : \Omega \to [0,\infty]$ on $(\Omega, \mathscr{F}, \mathbb{F})$ is called a **stopping time** (with respect to $\mathbb{F}$) if $\{\tau \leq t\} \in \mathscr{F}_t$ for all $t \geq 0$.*

Definition 2.13. *An adapted process $X = (X_t)_{t\geq 0}$ is said to be a **local martingale** if there exists a sequence (τ_n) of stopping times such that*

1. *The sequence is increasing and $\lim_{n\to\infty} \tau_n = \infty$, a.s.,*
2. *For every n, the stopped process $X^{\tau_n} = X_{\tau_n \wedge t}$ is a $\mathbb{F}$-martingale.*

The largest class of stochastic processes with respect to which the Ito integral, which we describe below, can be defined is the class of *semimartingales*. We restrict our attention to continuous semimartingales in this dissertation, and give the definition accordingly.

Definition 2.14. *An adapted, continuous, integrable process $X = (X_t)_{t\geq 0}$ is referred to as a **semimartingale** if it can be decomposed as $X_t = A_t + M_t$ for each t where M is an adapted, continuous local martingale and A is an adapted, continuous process of bounded variation.*

2.2 Stochastic Calculus

The single most important stochastic process in all of continuous-time finance is arguably Brownian motion. It belongs to the class of Levy processes, where it is uniquely identified as the only (stochastic) process which has (up to a version) continuous sample paths. Standard Brownian motion is both a martingale and a Markov process.

Definition 2.15. *Let $(\Omega, \mathscr{F}, \mathbb{F}, \mathscr{P})$ be a filtered probability space. An $\mathbb{F}$-adapted real-valued process B is called a standard $\mathbb{F}$-Brownian motion if*

1. $\mathbb{P}(B_0 = 0) = 1$,

2. *For every* $0 \le s < t < \infty$ *we have that* $B_t - B_s$ *is independent of* $\mathscr{F}_s$,

3. *For every* $0 \le s < t < \infty$ *we have that* $B_t - B_s \sim \mathscr{N}(0, t-s)$.

The sample paths of Brownian motion are in the standard sense nowhere differentiable, almost surely. As for integration with respect to Brownian motion, we come to one of the central constructs in stochastic calculus which is the Ito integral. To establish it in general terms we need to start with a special case.

Definition 2.16. *A process* $X \in \mathscr{L}_2 = \mathscr{L}_2(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ *is called* ***simple*** *if there exists a countable partition* $\Pi : 0 = t_0 < \ldots < t_n < \ldots$ *with* $\lim_{n\to\infty} t_n = \infty$ *such that* $X_t(\omega) = X_{t_j}(\omega)$ *for all* $t \in [t_j, t_{j+1})$, $j = 0, 1, 2, \ldots$ *and for all* $\omega \in \Omega$. *The subspace of square integrable simple processes is denoted by* $\mathscr{L}_2^0$.

We can define the Ito integral for simple processes in intuitive terms.

Definition 2.17. *Let* $X = (X_t)_{t\ge 0}$ *be a simple process. We define the* ***Ito integral*** *of* X *with respect the standard Brownian motion process* $B = (B_t)_{t\ge 0}$ *as*

$$I_t(X(\omega)) = \sum_{0\le j\le m-1} X_{t_j}(\omega)(B_{t_{j+1}}(\omega) - B_{t_j}(\omega)) + X_{t_m}(\omega)(B_t(\omega) - B_{t_m}(\omega)), \quad (2.1)$$

where $m = \max\{j : t_j \le t\}$

Adapted simple processes can be shown to be progressively measurable. Moreover, they can be used to approximate more general progressively measurable processes as the following result indicates.

Theorem 2.18. *Given any progressively measurable process* $X \in \mathscr{L}_2$, *there exists a sequence of simple processes* $X^{(n)} \in \mathscr{L}_2^0$ *such that*

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^t (X_s^{(n)} - X_s)^2 ds\right] = 0. \quad (2.2)$$

With the above in mind, we see that the definition of an integral for a simple process is extendable to more general processes.

Definition 2.19. *Let* $X = (X_t)_{t\ge 0} \in \mathscr{L}_2$ *be progressively measurable. The Ito integral of* X *with respect to the standard Brownian motion* $B = (B_t)_{t\ge 0}$ *is the unique, square integrable martingale* $I(X) = (I_t(X))_{t\ge 0}$ *which, for every* $t \ge 0$, *satisfies* $\lim_{n\to\infty} \mathbb{E}[(I_t(X^{(n)}) - I_t(X))^2] = 0$ *for every sequence* $\{X^{(n)}\}_{n=1}^{\infty} \subset \mathscr{L}_2^0$ *which satisfies* (2.2). *We write*

$$I_t(X) = \int_0^t X_s dB_s; \quad 0 \le t < \infty. \quad (2.3)$$

The following result, usually referred to as the Cameron-Martin-Girsanov theorem, shows that changing the probability measure when the filtration is generated by Brownian motion is particularly mathematically tractable. A Brownian motion under one measure stays the same under any other equivalent measure apart from a drift change.

Theorem 2.20. *Let $\lambda = (\lambda_t)_{t\geq 0}$ be a $\mathscr{F}$-predictable process such that*

$$\mathbb{E}^{\mathbb{P}}\left[\exp\left(\frac{1}{2}\int_0^T \lambda_t^2 dt\right)\right] < \infty,$$

and $B = (B_t)_{t\geq 0}$ be standard Brownian motion on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$. Then there exists a measure $\mathbb{Q}$ such that

1. *$\mathbb{Q}$ is equivalent to $\mathbb{P}$,*
2. *$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left[-\int_0^T \lambda_t dB_t - \frac{1}{2}\int_0^T \lambda_t^2 dt\right]$,*
3. *The process $\widetilde{B} = (\widetilde{B}_t)_{t\geq 0}$ defined as*

$$\widetilde{B_t} = B_t + \int_0^t \lambda_s ds$$

is a Brownian motion with respect to $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{Q})$.

In finance, the Radon-Nikodym derivative has the interpretation of being a state-price density, i.e., a stochastic process that prices every cash flow in the market. As such it has to be a strictly positive martingale, to prevent the presence of arbitrage. Exponential Brownian motion as it appears in the theorem above is precisely the process that satisfies this condition.

The processes considered throughout this thesis are generally presented as solutions to stochastic differential equations, which leads to a diffusion process more general than Brownian motion.

Definition 2.21. *Let B be a standard Brownian motion on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, and $\sigma = (\sigma_t)_{t\geq 0}$ be a progressively measurable process that satisfies the square-integrability condition*

$$\mathbb{E}\left[\int_0^t \sigma_s^2 ds\right] < \infty \text{ for each } t \geq 0,$$

and $\mu = (\mu_t)_{t\geq 0}$ be another progressively measurable process that satisfies the following integrability condition

$$\int_0^t |\mu_s| ds < \infty \text{ a.s. for each } t \geq 0.$$

*An **Ito process** is a stochastic process of the form*

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s.$$

With Ito processes formally introduced, we recall the well known Ito's lemma which plays the role of the fundamental theorem of calculus for stochastic integrals.

Definition 2.22. *Let X be an Ito process and $f = f(t,x)$ be a function of two variables with continuous partial derivatives, f_t, f_x and f_{xx}. Then, for every $T \geq 0$,*

$$f(T, X_t) = f(0, X_0) + \int_0^T \left(f_t(t, X_t)dt + \mu_t f_x(t, X_t) + \frac{1}{2}\sigma_t^2 f_{xx}(t, X_t) \right) dt + \int_0^T f_x(t, X_t)\sigma_t dB_t.$$

Finally, the following result tells us that given certain technical conditions on the drift μ and volatility σ the corresponding SDE admits a unique solution, which is a strong Markov process with continuous sample paths. In the following theorem, we write μ_t more explicitly as $\mu(t,x)$ to emphasize that it can be considered as a function of both time and the process X itself, with the same applying to σ_t and $\sigma(t,x)$.

Theorem 2.23. *Let $\mu : [0,T] \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,T] \times \mathbb{R}^n \to \mathbb{R}^{n\times m}$ with $T > 0$ be two progressively measurable functions satisfying*

$$|\mu(t,x)| + |\sigma(t,x)| \leq C(1 + |x|),$$

for some constant C and $x \in \mathbb{R}^n$, $t \in [0,T]$, and such that

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq D|x - y|,$$

for some constant D. Let Z be a square integrable random variable which is independent of the σ-algebra $\mathscr{F}^{(m)}$ generated by B_s, $s \geq 0$. Then the stochastic differential equation

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t, \quad 0 \leq t \leq T$$

with $X_0 = Z$ has a unique, square-integrable, pathwise-continuous solution X_t that is adapted to the filtration $\mathscr{F}_t^Z$ generated by Z and B.

2.2.1 The Partial Differential Equation Link

As we explain in greater detail in the next section, computing the price of a financial derivative is equivalent to computing its expected payoff under a given (subjective) probability measure. In some cases, probabilistic methods fail to provide us with computationally efficient means of calculating this expectation. Instead, we can analyze the problem from a deterministic perspective by taking advantage of the connection between expectations involving diffusion processes (i.e., Ito processes) on one hand, and partial differential equations on the other hand. We have the following result, known as the Feynman-Kac theorem.

Theorem 2.24. *A solution $f \in C^{2,1}(\mathbb{R}^n \times [0,T])$ to the partial differential equation (PDE)*

$$f_t(x,t)+\frac{1}{2}tr\Big(\sigma(x,t)\sigma(x,t)^\intercal f_{xx}(x,t)\Big)+f_x(x,t)\mu(x,t)-r(x,t)f(x,t)+h(x,t)=0, \tag{2.4}$$

subject to the terminal condition

$$f(x,T)=F(x), \tag{2.5}$$

where $x \in \mathbb{R}^n$, $t \in [0,T)$ is given by

$$\begin{aligned} f(x,t) =& \mathbb{E}^{\mathbb{Q}}\Big[\exp\Big\{-\int_t^T r(X_u,u)du\Big\}F(X_T)|\mathscr{F}_t\Big] \\ &+\mathbb{E}^{\mathbb{Q}}\Big[\exp\Big\{-\int_t^T r(X_u,u)du\Big\}h(X_s,s)ds|\mathscr{F}_t\Big], \end{aligned} \tag{2.6}$$

where X is given by Definition 2.21, and $X_t = x$.

The theorem tells us that we can solve the derivative pricing problem either by computing an expectation or by solving a PDE. While most PDEs do not have known closed-form solutions, finite difference methods provide a standard solution approach for the corresponding PDEs of many derivative pricing models in finance see e.g. Duffy [2013] for an overview.

2.2.2 Selected Stochastic Processes

Definition 2.25. *An adapted process $X = (X_t)_{t\geq 0}$ is referred to as geometric Brownian motion (GBM) if it follows the dynamics given by*

$$X_t = \mu\int_0^s X_s ds + \sigma\int_0^t X_s dB_s.$$

Definition 2.26. *An adapted stochastic process $X = (X_t)_{t\geq 0}$ is referred to as a Cox-Ingersoll-Ross (CIR) process, or a square root process, if it follows the dynamics given by*

$$X_t = \int_0^t \kappa(\theta - X_s)ds + \int_0^t \sigma\sqrt{X_s}dB_s.$$

It is said to satisfy the Feller condition if $2\kappa\theta > \sigma^2$.

Theorem 2.27. *A CIR process $X = (X_t)_{t\geq 0}$ that satisfies the Feller condition is strictly positive.*

Definition 2.28. *An adapted stochastic process $X = (X_t)_{t\geq 0}$ is referred to as a 3/2 process if it follows the dynamics given by*

$$X_t = \int_0^t \kappa X_s(\theta - X_s)ds + \int_0^t \sigma X_s^{3/2}dB_s.$$

Finally, Ito's lemma gives us the following relationship between the CIR process and the 3/2 process.

Theorem 2.29. *Let $X = (X_t)_{t\geq 0}$ be a CIR process, and define $Y_t = X_t^{-1}$ for all $t \geq 0$. Then we have that*

$$dY_t = \widetilde{\kappa}Y_t(\widetilde{\theta} - Y_t)dt + \widetilde{\sigma}Y_t^{3/2}dB_t, \tag{2.7}$$

where

$$\widetilde{\kappa} = \kappa\theta - \sigma^2, \quad \widetilde{\theta} = \frac{\kappa}{\kappa\theta - \sigma^2}, \quad \widetilde{\sigma} = -\sigma. \tag{2.8}$$

2.3 No-Arbitrage Pricing

Given a filtered probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, we think of a **market** as a collection of $J \in \mathbb{N}$ **assets**, where each asset is associated with an adapted stochastic process we refer to as its **spot price process**, along with a linear functional π which we refer to as a **pricing function**.

Let X be a spot price process. Given $t, T \in \mathbb{R}$ such that $0 \leq t \leq T$ and a function $\Psi : \mathscr{F}_T \to \mathbb{R}$, we denote by $\pi_t(\Psi(X,T))$ the **market price** of the **contingent claim** $\Psi(X,T)$ at time t. Naturally, we can also think of the spot price process of an asset in the market as a contingent claim where Ψ is the identity function.

An example of non-trivial contingent claims are so-called plain European call and put options with strike price K and maturity T, defined respectively by

$$\Psi_{EC}(X,K,T) = |X_T - K|^+,$$
$$\Psi_{EP}(X,K,T) = |K - X_T|^+,$$

where $|x|^+$ is shorthand for $\max\{x,0\}$. From now on we will use $\Psi(X,T), \Psi_T$ and $\Psi(X,T,K)$ interchangeably to denote a contingent claim with maturity T depending on the context.

Generally, participation in financial markets is voluntary. Implicit in the calculation of the price of a contingent claim is that its expected return is compared to the outcome of not investing in it to begin with. If the market offers an asset that yields a deterministic rate of return with no risk (i.e., a *risk-free asset*) then a profit seeking investor is expected to choose investing in this risk-free asset rather than simply leaving her capital inactive. This means the payoff of the contingent claim should be computed "in excess" of the return which the investor could collect with no risk. An equivalent way of stating this is that the future payoff of the contingent claim should be discounted by the payoff that the risk-free asset yields. The rate of return of the risk-free asset is referred to hereafter as the *risk-free rate*. Given a risk-free rate r and maturity T, the well known compound interest discount formula in discrete time finance extends naturally as a limit in continuous time as $\rho_T = e^{-rT}$. More generally, we can assume that the rate is stochastic, and rewrite the expression as $\rho_T = \exp\left(-\int_0^T r_s ds\right)$.

We now turn our attention to the pricing function π. Given certain conditions in the market, which we elaborate on below, the market price of a contingent claim can be computed as its risk-adjusted expected payoff

$$\pi_t(\Psi(X,T)) = \xi_t^{-1}\mathbb{E}^{\mathbb{P}}\left[\xi_T \Psi(X,T)|\mathscr{F}_t\right], \tag{2.9}$$

for some strictly positive stochastic process ξ sometimes referred to as a stochastic discount factor, a pricing kernel, or a state-price density. For the purpose of derivative pricing it proves more tractable to "factor out" ξ by changing to a subjective measure $\mathbb{Q}$ that absorbs ξ, obtaining

$$\pi_t(\Psi(X,T)) = \rho_t^{-1}\mathbb{E}^{\mathbb{Q}}\left[\rho_T \Psi(X,T)|\mathscr{F}_t\right]. \tag{2.10}$$

The measure $\mathbb{Q}$ is often referred to as the risk neutral measure, and the formulation above referred to as risk neutral pricing.

The aforementioned conditions that ensure that (2.10) is indeed a meaningful expression are technically involved. The fundamental assumption in asset pricing, and mathematical finance in general, is that financial markets

are arbitrage free. Informally, this means that a riskless profit cannot be made buying and selling the assets in the market that exceeds the return of the risk free asset (assuming it exists). This condition turns out to imply the existence of the equivalent martingale measure $\mathbb{Q}$, and vice versa. The theorem that establishes this equivalence is referred to as the First Fundamental Theorem of Asset Pricing (FFTAP). While it is easy to state and prove for a discrete state-space in discrete time, it is quite difficult in a fully continuous setting, where the concept of arbitrage becomes too narrow. In this sense the FFTAP for the general case becomes a collection of results with various, stronger definitions of the concept of arbitrage. We present the FFTAP in a "folk theorem" format, and refer to Downarowicz [2010] for a more comprehensive discussion.

Theorem 2.30. *A market is **essentially** arbitrage free if and only if there exists a martingale measure $\mathbb{Q}$ that is equivalent to the physical measure $\mathbb{P}$.*

The FFTAP does not imply that $\mathbb{Q}$ is unique. That issue is dealt with by the Second Fundamental Theorem of Asset Pricing (SFTAP). We begin with a couple of definitions.

Definition 2.31. *Let $X = (X_t^1, \ldots, X_t^J)_{t\in[0,T]}$ be the spot price process of the J assets in the market with an investment horizon $0 \leq T < \infty$, and let $Z = (Z_t^1, \ldots, Z_t^J)_{t\in[0,T]}$ be the discounted price process defined by $Z_t^j = X_t^j / X_0^j$ for $j = 1, \ldots, J$. Furthermore, let $L(Z)$ denote the set of all J-dimensional, predictable processes that are integrable with respect to the semi-martingale Z. A stochastic process $\lambda \in L(Z)$ is said to be an **admissible self-financing strategy** if*

1. *The discounted value process $V^*(\lambda) \equiv \sum_{j=1}^J \lambda^j Z^j$ is almost surely non-negative.*
2. *$V^*(\lambda)$ satisfies the **self-financing condition***

$$dV_t^*(\lambda) = \sum_{j=1}^J \lambda_s^j dZ_s^j, \quad t \in [0,T].$$

3. *$V^*(\lambda)$ is a martingale under $\mathbb{Q}$.*

Definition 2.32. *A contingent claim Ψ_T is said to be **attainable** if there exists an admissible self-financing trading strategy λ such that $V_T^*(\lambda) = \rho_T \Psi_T$. Furthermore, the market is said to be **complete** if every claim is attainable.*

In other words, we say that a market X is complete if every possible contingent claim in the market can be replicated by buying and selling traded assets in that market. We are now equipped to state the SFTAP.

Theorem 2.33. *Let $\mathbb{Q}$ be an equivalent martingale measure for the physical measure $\mathbb{P}$. The following statements are equivalent:*

1. *The market is complete under $\mathbb{Q}$,*
2. *$\mathbb{Q}$ is the only equivalent martingale measure for $\mathbb{P}$.*

An option pricing model usually comes with an assumption on the completeness of the market it describes. If all the stochastic factors in the model are traded then the model implies a complete market. An example of a model where this is the case is the Black Scholes model. It contains one stochastic factor that drives the price of the underlying stock which is, by definition, traded in the market. However, making the volatility of that model a stochastic process in itself[2] leads to a model that describes an incomplete market unless the volatility process itself is taken to be traded in the market.

We end this section with a very useful relationship between European put and call options that is known as *put-call parity* and is used extensively in the computational work behind the research results of this dissertation.

Theorem 2.34. *Let $P = \pi_t(\Psi_{EP}(X,K,T))$ and $C = \pi_t(\Psi_{EC}(X,K,T))$ denote the prices of a plain European put and call, respectively, with maturity T and strike price K with $0 \leq t \leq T$, and let r be the risk free rate of return. Then we have that $C + Ke^{-r(T-t)} = P + X_t$.*

This means once we have computed the price of a call option then we have with minimal computational effort the price of the corresponding put option as well, and vice versa.

2.4 Indifference Pricing

As discussed in the preceding section, no-arbitrage pricing considers the price process of the underlying assets in the market to be exogenous. When the market is complete, no mention of preferences of the investors in the market is necessary for reasoning about the price of its contingent claims. However, when the market is incomplete, an option pricing model requires an explicit assumption on the preferences over risk in the market.

The theory that explores this aspect of finance is referred to as general equilibrium asset pricing. The basic setup of this theory consists of a set C of

2. Here, we mean that the volatility becomes a stochastic process with its own stochastic factor. A volatility process that is only a deterministic function of the stochastic factor belonging to the underlying asset does not break the completeness of the market.

investment opportunities referred to as a consumption set, and a binary relation $\succsim$ over this set referred to as a preference relation. An investor whose preferences are represented by $\succsim$ prefers x over y, with $x, y \in C$ if and only if $x \succsim y$.

Definition 2.35. *The preference relation $\succsim$ is said to be **rational** if it possesses the following two properties:*

1. ***Completeness**: For all $x, y \in C$ we have that $x \succsim y$ or $y \succsim x$ (or both),*
2. ***Transitivity**: For all $x, y, z \in C$ if $x \succsim y$ and $y \succsim z$, then $x \succsim z$.*

A preference relation is by itself unwieldy when it comes to calculations. Instead, we hope to be able to find a tractable real-valued function which produces the same ordering.

Definition 2.36. *A function $u : C \to \mathbb{R}$ is called a **utility function** representing preference relation $\succsim$ if, for all $x, y \in C$ we have that $x \succsim y$ if and only if $u(x) \geq u(y)$.*

Although utility functions map elements in the consumption set to real numbers, they are not meant to be considered as cardinal functions, but as ordinal functions. This means that we consider two utility functions u and v as equivalent if there exists a monotonic transformation $g : \mathbb{R} \to \mathbb{R}$ such that $g(u) = v$ or vice versa.

The question now becomes whether for a given preference relation $\succsim$ we can always find a corresponding utility function that preserves the ordering of $\succsim$ over C. To answer that question, we must introduce one additional property.

Definition 2.37. *The preference relation $\succsim$ on C is **continuous** if it is preserved under limits. That is, for any sequence of pairs $\{(x_n, y_n)\}$ with $x_n \succsim y_n$ for all n, $x = \lim_{n\to\infty} x_n$ and $y = \lim_{n\to\infty} y_n$ we have that $x \succsim y$.*

We then have the following powerful result.

Theorem 2.38. *Let the preference relation $\succsim$ be rational and continuous. Then there exists a continuous utility function u that represents $\succsim$.*

The definitions and results presented so far apply to general consumption sets. Financial economics, however, deals with a very particular kind of consumption sets, where every good is simply a future cashflow. What distinguishes between these cashflows is the manner in which they are realized.

Hence, we need preferences that deal with risk. The most prominent type of such preferences in the literature is *expected utility*[3].

Definition 2.39. *Let C denote the consumption set consisting of the stochastic payoffs attainable by trading contingent claims in a given market. A utility function $U : C \to \mathbb{R}$ is of **expected utility** type if there exists a utility function $u : \mathbb{R} \to \mathbb{R}$ over deterministic outcomes such that for each $c \in C$*

$$U(c) = \mathbb{E}[u(c)].$$

The question now becomes when can preferences over risk be represented as expected utility. To answer that question we need the following definition.

Definition 2.40. *The preference relation $\succsim$ over the consumption set C satisfies the **independence axiom** if for all $c, c', c'' \in C$ and $\alpha \in (0,1)$ we have $c \succsim c'$ if and only if $\alpha c + (1-\alpha)c'' \succsim \alpha c' + (1-\alpha)c''$.*

In other words, if we mix two stochastic payoffs with a third one, then the preference ordering of the two resulting mixtures does not depend on the particular third stochastic payoff.

What follows is one of the most important results in the theory of choice under uncertainty.

Theorem 2.41. *If a rational preference relation $\succsim$ satisfies the axioms of continuity and independence then it is representable by a function of the expected utility type.*

One of the primary assumptions we make when modeling financial markets is that investors are risk averse, which explains why systemic risk is compensated in financial markets.

Definition 2.42. *A preference relation $\succsim$ implies **risk aversion** if for any stochastic payoff c it ranks the certain payoff $\mathbb{E}[c]$ higher in the preference order than c itself.*

One of the advantages of the expected utility format is that it allows us to express risk aversion in a particularly tractable way. The following theorem is a straightforward application of Jensen's inequality.

Theorem 2.43. *Let U be a utility function of expected utility type, with an associated utility function $u : \mathbb{R} \to \mathbb{R}$ over deterministic outcomes. The preferences represented by U imply risk aversion if and only if u is concave.*

3. Note that for the remainder of this subsection we drop explicit references to a filtered probability space in the context of the market made in previous subsections as the following definitions and results apply in a more general setting.

In financial economics, the concavity of the function u corresponds directly with how much the investor needs to be compensated in order to take on risk in the market. The greater the concavity, the more risk averse the investor, and thus the higher the compensation she requires.

According to general equilibrium theory, market prices form in a conjunction with an equilibrium between supply and demand. And within this framework, supply and demand in financial markets are determined by market participants who seek to trade in the market in an effort to maximize their expected return while at the same time minimizing the risk they must take on. This pursuit can be expressed in a simple way as a utility maximization problem.

Definition 2.44. *Consider an investor with expected utility preferences in a market consisting of J assets with a payoff process $\boldsymbol{G}_t = (G_{t,1}, \ldots, G_{t,J})$ and an associated price vector $\boldsymbol{\pi}_t = (\pi_{t,1}, \ldots, \pi_{t,J})$. Then the investor's (dynamic)* ***portfolio choice problem*** *is given recursively by*

$$\begin{aligned} &\max_{\boldsymbol{\phi}} \; u(c_t) + \beta \mathbb{E}_t\left[u(c_{t+1})\right] \\ &\text{s.t.} \quad c_t = \boldsymbol{\phi}_t^\top \boldsymbol{G}_t - \boldsymbol{\pi}_t^\top \boldsymbol{\phi}_{t+1}, \\ &\qquad c_{t+1} = \boldsymbol{\phi}_{t+1}^\top \boldsymbol{G}_{t+1} - \boldsymbol{\pi}_{t+1}^\top \boldsymbol{\phi}_{t+2}, \end{aligned} \tag{2.11}$$

where $u : \mathbb{R} \to \mathbb{R}$ is the investor's utility function over deterministic outcomes, β is the investor's deterministic discount rate, and $\boldsymbol{\phi}_t = (\phi_{t,1}, \ldots, \phi_{t,J})$ is the investor's portfolio choice at time t.

When the aggregate demand of the market is assumed to have a sufficiently tractable form, the general equilibrium theory framework allows us to derive the stochastic discount factor for the market.

Definition 2.45. *A stochastic process ξ is referred to as a* ***stochastic discount factor*** *if for every asset with a payoff process G it holds that the price of the asset at time t is given by $\pi_t = \xi_t^{-1} \mathbb{E}_t[\xi_{t+1} G_{t+1}]$ for all t.*

Theorem 2.46. *If the representative investor's portfolio choice problem is given by* (2.11)*, with u differentiable, then the stochastic discount factor is given by*

$$\xi_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}.$$

The economic interpretation of the stochastic discount factor is that it gives the relative value of a standard unit of capital for each possible future state of the market. The intuition behind this statement is that in a realized future state in which the payoff is large across assets in the market, the value of a

single unit of capital is less than in a realized future state where the overall payoff is smaller. Therefore, contingent claims that are negatively correlated with the market portfolio are priced higher than contingent claims that are positively correlated with the market.

2.4.1 Selected Utility Functions

We begin with the notions of absolute and relative risk aversion

Definition 2.47. *Given any $u \in C^2(l,\infty)$, the corresponding **coefficient of absolute risk aversion** $A^u : (l,\infty) \to \mathbb{R}$ and **coefficient of relative risk aversion** $R^u : (l,\infty) \to \mathbb{R}$ are defined by*

$$A^u(c) = -\frac{u''(c)}{u'(c)} \text{ and } R^u(c) = cA^u(c).$$

In the definition above, l is present in case u is not defined for negative wealth. We now give the definition of one of the most frequently used classes of utility functions over deterministic outcomes in financial economics.

Definition 2.48. *The utility function u is of **hyperbolic absolute risk aversion (HARA)** type with coefficients $(\alpha,\beta) \in \mathbb{R}^2$ if it belongs to $C^2\left(\{c : \alpha + \beta c > 0\}\right)$ and*

$$A^u(c) = \frac{1}{\alpha + \beta c}, \quad \alpha + \beta c > 0.$$

The two most prominent elements in the HARA class are retrieved when either $A^u(c) = \gamma$ which for $\gamma > 0$ corresponds to setting $(\alpha,\beta) = (1/\gamma, 0)$, or when $R^u(c) = \gamma$ which for $\gamma > 0$ corresponds to setting $(\alpha,\beta) = (0, 1/\gamma)$.

Definition 2.49. *The utility function given by*

$$u(c) = \begin{cases} \dfrac{1 - e^{-\gamma c}}{\gamma} & \text{if } \gamma \neq 0, \\ c & \text{if } \gamma = 0, \end{cases}$$

*is referred to as the **exponential utility function**. It is the unique function that exhibits constant absolute risk aversion (CARA) and is therefore also referred to as the **CARA utility function**.*

Definition 2.50. *The utility function given by*

$$u(c) = \begin{cases} \dfrac{c^{1-\gamma} - 1}{1 - \gamma} & \text{if } \gamma \neq 1, \\ \ln(c) & \text{if } \gamma = 1, \end{cases}$$

*is referred to as the **power utility function**. It is the unique function that exhibits constant relative risk aversion (CRRA), and is therefore also referred to as the **CRRA utility function**.*

2.5 Computational Finance

Here we list some results that have proven fundamental with respect to the computational aspect of option pricing.

2.5.1 Monte Carlo Pricing

As discussed in 2.3, the price of an option can be computed as the discounted expectation of its payoff at maturity under the risk neutral measure. Often this expectation cannot be derived analytically. Transforming the expectation problem to a PDE problem which we can solve using finite difference methods is one approach. Another is computing the expectation through simulation. The method, which is generally referred to as the Monte Carlo method, consists of sampling the probability dynamics of the option payoff, and computing the sample average which, under fairly general conditions, converges towards the true value of the expectation. More specifically, for a contingent claim $\Psi(X,T)$ on the filtered probability space $(\Omega,\mathscr{F},\mathbb{F},\mathbb{Q})$ we compute the following estimate $\overline{\psi}^{(N)}$ of $\mathbb{E}^{\mathbb{Q}}[\Psi(X,T)]$:

$$\overline{\psi}^{(N)} = \frac{1}{N}\sum_{i=0}^{N}\psi_i, \tag{2.12}$$

where the ψ_i are realizations of the i.i.d. random variables Ψ_i with $\mathbb{E}^{\mathbb{Q}}[\Psi_i] = \mathbb{E}^{\mathbb{Q}}[\Psi(X,T)]$ for $i = 1,\dots,N$.

For most situations where Monte Carlo methods are a computationally competitive approach, we need to simulate the entire path $X(\omega_i)$ for the randomly drawn ω_i, to obtain the realized payoff of the claim $\Psi(X(\omega_i),T)$ for each $i = 1,\dots,N$. A prototypical method for achieving this is the Euler-Maruyama (EM) scheme. Let X be an Ito process with instantaneous drift and volatility functions μ and σ respectively, that are Lipschitz continuous. Consider a partition $\boldsymbol{\tau} = [0,\Delta t,2\Delta t,\dots,N\Delta t]$ of the interval $[0,T]$ with $N\Delta t = T$. An EM approximation to X on the interval $[0,T]$ using $\boldsymbol{\tau}$ with initial value X_0 is given by

$$\hat{X}_{(n+1)\Delta t} = \mu(n\Delta t,\hat{X}_{n\Delta t})\Delta t + \sigma(n\Delta t,\hat{X}_{n\Delta t})\Delta B_n,$$

where $\Delta B_n = B_{n+1} - B_n \sim N(0,\Delta t)$. Since $\hat{X}$ is an approximation to X we need to consider the error involved, which we can reason about in formal terms using the convergence properties of the scheme.

Definition 2.51. *We say that the numerical scheme producing $\hat{X}$ is **strongly convergent** if*

$$\lim_{\Delta t\to 0}\mathbb{E}[|X_T - \hat{X}_T|] = 0.$$

Furthermore, if there exists a constant K_T such that

$$\mathbb{E}\left(\left|X_T - \hat{X}_T\right|\right) \leq K_T(\Delta t)^\gamma,$$

*then we say that the scheme is **strongly convergent with order** γ.*

Definition 2.52. *We say that the numerical scheme producing $\hat{X}$ is **weakly convergent** if*

$$\lim_{\Delta t \to 0} \left|\mathbb{E}[g(X_T)] - \mathbb{E}[g(\hat{X}_T)]\right| = 0,$$

for every polynomial g. Furthermore, if there exists a constant K_T^g such that

$$\left|\mathbb{E}[g(X_T) - \mathbb{E}[g(\hat{X}_T)]\right| \leq K_T^g(\Delta t)^\gamma,$$

*we say that the scheme is **weakly convergent with order** γ.*

We can think of the weak convergence rate as a measurement of the error in the expected payoff, and the strong convergence rate as the error per path.

Theorem 2.53. *The EM scheme is strongly convergent with order $\frac{1}{2}$ and weakly convergent with order* 1.

A strong convergence rate of $\frac{1}{2}$ means the EM scheme has fairly poor pathwise convergence properties. A refinement of the EM scheme was developed by Milstein [1975], and is given by

$$\begin{aligned}\hat{X}_{(n+1)\Delta t} =& \mu(n\Delta t, \hat{X}_{n\Delta t})\Delta t + \sigma(n\Delta t, \hat{X}_{n\Delta t})\Delta B_n \\ &+ \frac{1}{2}\sigma'(n\Delta t, \hat{X}_n)\sigma(n\Delta t, \hat{X}_n)((\Delta B_n)^2 - \Delta t),\end{aligned}$$

where σ' is the first derivative of σ with respect to its second argument. We have the following convergence results.

Theorem 2.54. *The Milstein scheme is strongly convergent with order* 1 *and weakly convergent with order* 1.

From a statistics perspective, the estimator $\overline{\psi}^{(N)}$ is by itself not particularly efficient. Assuming $\Psi_1, \ldots, \Psi_N$ are i.i.d. with $\mathrm{Var}[\Psi_i] = \sigma_\psi^2$, the Central Limit Theorem implies that

$$\sqrt{N}\left(\overline{\Psi}^{(N)} - \mathbb{E}[\Psi(X,T)]\right) \xrightarrow{d} N(0, \sigma_\psi^2),$$

which implies that with a sample size of N the standard deviation of the estimate is only shrinking at a rate of $\frac{1}{\sqrt{N}}$. Many strategies have been developed to increase the efficiency of Monte Carlo estimators in finance. A

comprehensive overview of these can be found in chapter 4 in Glasserman [2013]. One of these in particular deserves mention as its broad applicability makes it ubiquitous in the field of Monte Carlo pricing, and is utilized in the computational implementation of the research on the weighted Monte Carlo method presented later in this dissertation. It is referred to as the antithetic variates method.

Let $\Psi_1, \ldots, \Psi_N$ be i.i.d. such that $\mathbb{E}[\Psi_i] = \mathbb{E}[\Psi(X,T)]$ for every $i = 1, \ldots, N$ as before. Assume now we have a second i.i.d. sequence, $\hat{\Psi}_1, \ldots, \hat{\Psi}_N$ that also satisfies $\mathbb{E}[\hat{\Psi}_i] = \mathbb{E}[\Psi(X,T)]$ for every $i = 1, \ldots, N$, but where Ψ_i and $\hat{\Psi}_i$ are not necessarily independent. We have that an unbiased estimator for $\mathbb{E}[\Psi(X,T)]$ is given by

$$\frac{1}{2N}\sum_{i=1}^{N}\Psi_i + \frac{1}{2N}\sum_{i=1}^{N}\hat{\Psi}_i. \tag{2.13}$$

The basic formula for the variance of the sum of two random variables gives us that

$$\mathrm{Var}[\Psi_i + \hat{\Psi}_i] = \mathrm{Var}[\Psi_i] + \mathrm{Var}[\hat{\Psi}_i] + 2\mathrm{Cov}[\Psi_i, \hat{\Psi}_i],$$

which implies that $\mathrm{Var}[\Psi_i + \hat{\Psi}_i]$, and hence of the estimator (2.13) is smaller when Ψ_i and $\hat{\Psi}_i$ are negatively correlated. If we use, for example, the Inverse Transform Method to sample the distribution we can for every realization u_i of the uniformly distributed random variable $U \sim \mathrm{Uniform}[0,1]$ calculate the antithetic realization of ψ_i using $1 - u_i$. This not only gives us a more efficient estimator as implied by (2.13), but also reduces the computational cost of the simulation because we now get two realizations for the price of one[4].

2.5.2 Pricing With Characteristic Functions

Modern no-arbitrage pricing models show tremendous flexibility in terms of the empirical features in financial markets they can capture. However, most of the dynamics of which we can perceive in this regard are unlikely to yield a known transition density. This is a significant drawback since it means the computational cost of solving the model can become prohibitive, particularly when the model is being calibrated or undergoing sensitivity analysis, or when the model consists of a large number of factors. In addition, not having any grasp of the underlying distribution is a hindrance in terms of the theoretical exploration of these models.

4. While we still need to compute $\hat{\psi}_i$, the fact is that randomly generating numbers on most computer systems is very slow.

While the set of models for which we have an explicit transition density is very small, the set of models for which we know the characteristic function of the density is much larger. Through that we are able to unlock a great deal of tractability, often resulting in closed-form solutions (up to Fourier inversion) for certain derivative prices, while for others it gives us good approximations that can be used to speed up less computationally efficient methods such as Monte Carlo simulations.

The characteristic function of a random variable X with an associated probability measure $\mathbb{Q}$ is defined as

$$\phi_X(u) = \mathbb{E}^{\mathbb{Q}}[e^{iuX}] = \int_{\Omega} e^{iuX} d\mathbb{Q}, \tag{2.14}$$

and the following basic properties apply to it:

1. $\phi_X(u)$ always exists,
2. $\phi_X(u)$ is a continuous function in $u \in \mathbb{R}$,
3. $\phi_X(0) = 1$ for any distribution,
4. $\left|\phi_X(u)\right| \leq 1$ for all u,
5. $\overline{\phi_X(u)} = \phi_X(-u)$,
6. If $Y = a + bX$ then the characteristic function for Y is $e^{iua}\phi_X(bu)$,
7. If X and Y are stochastically independent with characteristic functions ϕ_X and ϕ_Y, respectively, then the characteristic function of $Z = X + Y$ is given by $\phi_X(u)\phi_Y(u)$,
8. The kth moment of X, assuming it exists, is given by

$$\mathbb{E}\left[X^k\right] = \frac{1}{i^k} \frac{d^k \phi_X(u)}{du^k}\Big|_{u=0}$$

Most of these properties derive directly from the properties of the Fourier transform since the characteristic function can simply be interpreted as the Fourier transform of the probability density of X (assuming it exists).

The fundamental property of characteristic functions is the one-to-one mapping between them and the set of probability measures. Definition (2.14) shows us how we go from a random variable X with an associated probability measure $\mathbb{Q}$ to its corresponding characteristic function $\phi_X(u)$. As for going in the opposite direction, we have the following theorem.

Theorem 2.55. *Let ϕ_X be the characteristic function of a random variable X with probability distribution given by the measure $\mathbb{Q}$ on $\mathscr{B}(\mathbb{R})$. If $a < b$, then*

$$\mathbb{Q}((a,b)) + \frac{\mathbb{Q}(\{a\}) + \mathbb{Q}(\{b\})}{2} = \lim \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-iua} - e^{-iub}}{iu} \phi_X(u)\, du.$$

As we can see, if $\{a\}$ and $\{b\}$ are null sets with respect to $\mathbb{Q}$ then the formula above simply retrieves $\mathbb{Q}$. In simpler terms, we get the cumulative distribution function as

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-iux} \phi_X(u)}{iu}\, du,$$

and the probability density function as

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \phi_X(u)\, du.$$

The importance of the inversion theorem in the context of option pricing is that we can use it to numerically extract the probability distribution of contingent claims such as options which in return enables us to compute their prices in an efficient way.

We conclude this subsection by mentioning the class of affine models, which make up a large part of the class of models for which we have a characteristic function in closed form. It includes both the Black-Scholes model as well as the Heston model of stochastic volatility both of which we use in the last chapter of this dissertation.

Definition 2.56. *A stochastic process $X = (X_t)_{t \geq 0}$ is called **affine** if the covariance matrix $\sigma(x)\sigma(x)^\top$ and the drift $\mu(x)$ are affine in x. That is,*

$$\sigma(x)\sigma(x)^\top = a + \sum_{i=1}^{d} x_i \alpha_i,$$

$$\mu(x) = b + \sum_{i=1}^{d} x_i \beta_i,$$

for some real valued $d \times d$ matrices a and α_i and real valued d-vectors b and β_i.

The following result shows that given an affine process X, the problem of finding its characteristic function boils down to solving a system of ordinary differential equations.

Theorem 2.57. *Let $X = (X_t)_{t\geq 0}$ be an affine process with drift $\mu(x)$ and covariance matrix $\sigma(x)\sigma(x)^\top$. Then the characteristic function $\phi_{X_T|\mathscr{F}_t}(u)$ of X_T conditioned on $\mathscr{F}_t$ is exponential-affine in X_t for all $t \leq T$. That is, there exist $\mathbb{C}$-and $\mathbb{C}^d$-valued functions $\xi(t,u)$ and $\psi(t,u)$ with jointly continuous t-derivatives, such that*

$$\mathbb{E}\left[e^{iu^\top X_T|\mathscr{F}_t}\right] = e^{\xi(T-t,u)+\psi(T-t,u)^\top X_t},$$

for all $u \in \mathbb{R}^d$ and $t \leq T$. Moreover, ξ and $\psi = \left[\psi^{(1)},\ldots,\psi^{(d)}\right]^\top$ solve the system of Riccati equations

$$\begin{aligned}\partial_t \xi(t,u) &= \frac{1}{2}\psi(t,u)^\top a\psi(t,u) + b^\top\psi(t,u),\\ \partial_t \psi_i(t,u) &= \frac{1}{2}\psi(t,u)^\top \alpha_i\psi(t,u) + \beta_i^\top\psi(t,u),\ 1\leq i\leq d,\end{aligned}$$

subject to the boundary conditions $\xi(0,u) = 0$ and $\psi(0,u) = iu$.

As we can see in the preceding theorem, ξ can be computed by simple integration once we know ψ. More specifically, we have that

$$\xi(t,u) = \int_0^t \left(\frac{1}{2}\psi(s,u)^\top a\psi(s,u) + b^\top\psi(s,u)\right) ds.$$

Chapter 3

On the Calibration of the 3/2 Model

3.1 Introduction

Since the Black-Scholes model's rise to prominence in the field of option pricing, numerous extensions to it have been proposed in an attempt to address its well known limitations. One category of such extensions involves generalizing the Black-Scholes by making the volatility of the underlying asset(s) a stochastic process. While the assumption of stochastic volatility has been shown to enable option pricing models to reproduce many of the empirical features of financial markets which the Black Scholes model misses, the added model complexity brings about its own challenges. In particular, they generally lack a closed-form solution. This becomes particularly significant during the calibration of the model to market data, where a potentially large set of benchmark options need to be priced by the model up to several thousand times, making the computational tractability of the model a key issue.

While stochastic volatility models generally do not yield themselves to a simple solution in the same way as the Black Scholes model, several special cases do admit a density for which we know the characteristic function in closed form. This essentially reduces the computation of an option price to a Fourier inversion, which in most cases is far cheaper computationally than solving the model through either a Monte Carlo simulation or a finite difference scheme.

The best known models with this property are those which correspond to an (exponentially) affine characteristic function, one example being the square

root volatility model of Heston [1993] and its many (affine) extensions (see e.g. Bates [1996]; Sepp [2008]; Wong and Lo [2009]). However, several studies done on the S&P 500 index seem to suggest that the dynamics of the underlying process are not affine, see e.g. Poteshman [1998]. In particular, Bakshi et al. [2006] use the VIX index to estimate several specifications of the volatility process, including

$$d v_t = \left(\alpha_0 + \alpha_1 v_t + \alpha_2 v_t^2 + \frac{\alpha_3}{v_t} \right) dt + \beta_0 v_t^{\beta_1} dW_t, \tag{3.1}$$

with $0 \leq t \leq T$ for a finite horizon T. Their findings suggest that both α_2 and α_3 are significant, and that β_1 is around 1.27.

A stochastic volatility model with the potential to better match these empirical findings (Carr and Sun [2007]) is the so-called 3/2 model. It is given by

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_t} S_t dB_t, \\ dV_t &= \kappa(\eta V_t - V_t^2) dt + \sigma {V_t}^{3/2} dW_t, \end{aligned} \tag{3.2}$$

where B and W are standard Brownian motions with $\text{Cov}(dB_t, dW_t) = \rho dt$. Although non-affine, the characteristic function for the corresponding density is known in closed form, implying the same tractability as we have for affine models.

A comparison of the 3/2 model and the Heston model reveals several fundamental differences. The 3/2 model allows for much stronger and more rapid volatility deviations away from the zero boundary than the Heston model. In addition, while the dynamics of the Heston model predict that the implied volatility skew flattens when the instantaneous volatility increases, the dynamics of the 3/2 model predict the opposite behavior. As such, the 3/2 model fits the empirical assumptions of risk managers better than the Heston model (Drimus [2012]).[5]

A potential issue, however, is that the characteristic function is considerably more complex than that of the Heston model. In particular, it involves the confluent hypergeometric function, $M(\alpha, \beta, z)$ which is defined as the solution to

$$z \frac{d^2 w}{dz^2} + (\beta - z) \frac{dw}{dz} - \alpha w = 0. \tag{3.3}$$

This nests a wide variety of special functions, as well as the exponential function, which is the most computationally expensive function in the char-

5. Moreover, we do not have to choose between them. Knowing the characteristic function for both of these models means we can combine them into one multi-factor model for which we also know the characteristic function in closed form (Grasselli [2017]). This is a consequence of the convolution property of Fourier transforms, see 3.4.3 for a more in-depth discussion.

acteristic function of the Heston model. The computational cost of evaluating it in the parameter space that corresponds to plausible values of the parameters of the 3/2 model can be up to two orders of magnitude greater than for the exponential function.[6]

For a one factor stochastic volatility specification, this puts the characteristic function approach to option pricing with the 3/2 model rather close to a finite difference scheme in terms of speed. While this might not be a problem for evaluating the price of a single option, it certainly has a big impact on the calibration of the model. For this reason, speeding up the calibration process is crucial to making the model computationally efficient enough for practitioners.

With this in mind, we propose a calibration method which uses the analytic gradient of the characteristic function. The compact form of the analytic gradient we derive allows us to develop a very efficient caching method for the partial derivatives, in addition to the strike and maturity dimensions. This approach turns out to be an order of magnitude faster in terms of computing the gradient of the objective function compared to a centered finite difference gradient approximation, with a similar speedup in the calibration of the model. The speedup factor is similar to what is reported in Cui et al. [2017], where the authors give a comprehensive exposition of the calibration of the Heston model, that includes the derivation of the analytic gradient of the characteristic function of the Heston model.

Lastly, as has been reported in the literature (see e.g. Cont and Tankov [2004]; Guillaume and Schoutens [2010]; Mikhailov and Nögel [2003]), the nonlinear least squares approach to calibrating option pricing models to market data is often accompanied by issues of non-convexity and numerical instability. More precisely, the objective function can exhibit several different local minima and using a purely gradient-based optimization can lead to very different parameter values depending on the initial guess. The precise choice of benchmark options also tends to have an impact on what values we obtain from the calibration procedure, and the optimal parameter values can fluctuate considerably between consecutive trading days.[7]

To tackle these issues for the 3/2 model, we propose using a Tikhonov type of regularization formulation of the least squares problem, where the regularization point is computed using MCMC estimation of the 3/2 model, and the damping matrix is the inverse covariance matrix of the parameter estimates.

6. This assumes a standard implementation of the confluent hypergeometric function for complex valued arguments, see Abramowitz and Stegun [1964]
7. Of course, we could also interpret this in a way that the model should incorporate parameters that are themselves stochastic processes, but that is outside the scope of this text.

Since we are only after the parameter values under the risk neutral measure, we propose using the volatility index VIX as a proxy for the instantaneous volatility which greatly simplifies the MCMC estimation procedure.

The remainder of this chapter is as follows. In section 3.2, we derive the analytic gradient for the characteristic function of the 3/2 model and present the caching algorithm for quickly evaluating it for a set of benchmark options. In section 3.3 we describe a regularized version of the calibration problem, and give the MCMC estimators we use for computing the regularization point and the damping matrix. Finally, in section 3.4 we present the results of our numerical experiments on these methods.

3.2 Calibration of the 3/2 Model with an Analytic Characteristic Function

Given a set of market quoted plain vanilla options, $i = 1,\dots,I$, we characterize option i as the pair $(K,T)_i$, where K is its strike price, and T is its maturity. We observe the market price π_i^{Market} for each of these options at time t, and wish to infer the parameters of our pricing model using these observed market prices. The standard approach to solving this problem involves solving the weighted (nonlinear) least squares minimization problem,

$$\boldsymbol{\Theta}^* = \underset{\boldsymbol{\Theta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta})^2 \right\}, \tag{3.4}$$

where

$$f_i(\boldsymbol{\Theta}) \equiv \pi(\boldsymbol{\Theta};(K,T)_i) - \pi_i^{\text{Market}}.$$

Here, w_i is a weight which determines the influence of option price i on the calibration, and $\pi(\boldsymbol{\Theta};(K,T)_i)$ is the price of option i given by the model, given the parameter values $\boldsymbol{\Theta}$. In our case, the vector of parameters is given by $\boldsymbol{\Theta} = \{\kappa,\eta,v,\sigma,\rho\}$. Note that while the spot volatility $v = V_t$ is strictly speaking not a parameter of the model, it is unobserved and therefore included as a decision variable in the optimization as well.

The set of quoted options we consider for our numerical studies are plain European options. More precisely, denoting with S_T the price of the underlying asset at time T, the price π^{call} of a European call option at time t with strike K and maturity T is given by $\pi^{\text{call}} = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left[|S_T - K|^+\right]$. Likewise, the price π^{put} of a European put option at time t with the same strike and maturity is given by $\pi^{\text{put}} = e^{-r(T-t)}\mathbb{E}_t^{\mathbb{Q}}\left[|K - S_T|^+\right]$. Here, the superscript $\mathbb{Q}$ refers to the fact that the expectation is taken under the risk-neutral measure, rather than the objective measure.

Since the probability density of S_T is not known in closed form for the 3/2 model, a straightforward calculation of the expectation of the payoff is not possible. On the other hand, we do know the characteristic function of the density of $\ln(S_T)$. More precisely, the characteristic function, $\phi(u) \equiv \mathbb{E}^{\mathbb{Q}}\left[e^{iuX_T}\right]$, where $X_T \equiv \ln(S_T)$, is known in closed form. With the parameter set $\boldsymbol{\Theta}$ explicitly accounted for, and letting $x = \ln(S_t)$, it is given by

$$\phi(\boldsymbol{\Theta}; u) = e^{iux}\frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\zeta^{\alpha}M\left(\alpha;\beta;-\zeta\right), \tag{3.5}$$

where $M(\alpha;\beta,-\zeta)$ is the confluent hypergeometric function of the first kind,

$$M(\alpha;\beta;-\zeta) \equiv \sum_{n=0}^{\infty}\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!}, \tag{3.6}$$

with

$$(x)_n \equiv \begin{cases} \prod_{i=0}^{n-1}(x+i) & \text{if } n>0, \\ 1 & \text{if } n=0, \end{cases}$$

and α, β, and ζ are given by

$$\begin{aligned} \zeta &\equiv \frac{2\kappa\eta}{v\sigma^2\left(e^{\kappa\eta(T-t)}-1\right)}, \\ \alpha &\equiv -\left(\frac{1}{2}-\frac{p}{\sigma^2}\right)+\sqrt{\left(\frac{1}{2}-\frac{p}{\sigma^2}\right)^2+\frac{2q}{\sigma^2}}, \\ \beta &\equiv 2\left[\alpha+1-\frac{p}{\sigma^2}\right], \\ p &\equiv -\kappa+i\sigma\rho u, \\ q &\equiv \frac{iu}{2}+\frac{u^2}{2}. \end{aligned} \tag{3.7}$$

Armed with the characteristic function, the next step is to apply an inverse Fourier transform to calculate the option price. The details of this step can vary depending on which technique we use, see e.g. Aichinger and Binder [2013]. We have selected the following formula by Attari [2004]:

$$\begin{aligned} &\pi^{\text{call}}(S_t,T,K) = S_t - \frac{1}{2}e^{-r(T-t)}K \\ &-\frac{e^{-r(T-t)}K}{\pi}\int_0^{\infty}\frac{\left(\text{Re}(\phi(u))+\frac{\text{Im}(\phi(u))}{u}\right)\cos(uk)+\left(\text{Im}(\phi(u))-\frac{\text{Re}(\phi(u))}{u}\right)\sin(uk)}{1+u^2}du, \end{aligned} \tag{3.8}$$

where $k = \ln(K)$. Comparing the speed of different Fourier based methods for pricing vanilla options, Kilin [2011] concludes that using adaptive

quadrature, with intermediary caching with the formula above, vastly outperforms the fast Fourier transform (FFT) method of Carr and Madan [1999], as well as the fractional FFT method of Chourdakis [2005]. We present an extension of this idea in the following section, which takes advantage of the special structure of the characteristic function of the 3/2 model.

3.2.1 The Analytic Gradient of the Characteristic Function

As mentioned in the introduction, the computation of the characteristic function of the 3/2 model is very expensive, which makes a gradient-based approach to finding the solution to (3.4) the only practical option. More specifically, methods which compute

$$\mathbf{\Theta}_{j+1} = \mathbf{\Theta}_j - \gamma_j H^{-1} \nabla_\Theta \left(\sum_{i=1}^{I} w_i f_i(\mathbf{\Theta}_j)^2 \right)$$

iteratively until one or more of the user provided stopping criteria are satisfied. Here γ_j is the step length, and H^{-1} is either the inverse Hessian of the objective function, or an approximation to the inverse Hessian. Given the existence of a feasible minimum, this iterative process should terminate for some j^* such that $\mathbf{\Theta}^* = \mathbf{\Theta}_{j^*}$. Calculating the full Hessian is very time consuming, so instead we utilize a Quasi-Newton method for solving (3.4), which is the standard approach to this problem. Quasi-Newton methods construct an approximation H^{-1} to the inverse Hessian by means of the gradient. This means that the gradient computation for (3.4) is the main computational work of finding $\mathbf{\Theta}^*$. How γ and H^{-1} are calculated depends on the specific Quasi-Newton method. For our calibration tests in section 3.4, we performed the minimization using the Levenberg-Marquardt algorithm (Marquardt [1963]) which is a typical approach to solving nonlinear least squares problems such as the calibration problem of (3.4).

Without an analytic form for the gradient, a finite difference approximation can be used, such as the simple forward difference quotient, or the symmetric difference quotient. These are given by

$$D_+ f(x) \equiv \frac{f(x+h) - f(x)}{h}, \quad D_\pm f(x) \equiv \frac{f(x+h) - f(x-h)}{2h},$$

respectively, with the latter being significantly more accurate than the first. There are, however, well known numerical issues that go along with using either of these for the purposes of optimization. To begin with, they are ill-conditioned and subject to cancellation errors when h gets small enough. On the other hand, putting h too large results in truncation errors. Even if

h is optimally chosen, the value of D_+ ($D_\pm$) is only half as accurate (two-thirds as accurate) in terms of significant digits as $f(x)$ (see e.g. Griewank and Walther [1987] for a more in-depth discussion).

Another issue is that if the optimization is constrained, which indeed it is in our case, invalid values can be generated on the boundary by the finite difference approximation due to the $x \pm h$ step. Lastly, the numerical gradient approach misses out on the interdependency between the different partial derivatives which can be used to speed up the computation of the gradient.

With that in mind, deriving and using the analytic gradient of the objective function (3.4) is well motivated in our attempt to speed up the calibration of the 3/2 model. We begin by noting that

$$\nabla_\Theta \left(\sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta})^2 \right) = 2 \sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta}) \nabla_\Theta f_i = 2 \sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta}) \nabla_\Theta \pi(\boldsymbol{\Theta}; (K,T)_i).$$

Furthermore, given the formula for the price $\pi(\boldsymbol{\Theta}; (K,T)_i)$ for $i = 1, \ldots, I$ above, we see that

$$\nabla_\Theta \pi(\boldsymbol{\Theta}; (K,T)_i) = -\frac{e^{-r(T-t)}K}{\pi} \times \int_0^\infty \frac{\left(\mathrm{Re}(\nabla_\Theta \phi(u)) + \frac{\mathrm{Im}(\nabla_\Theta \phi(u))}{u} \right) \cos(uk) + \left(\mathrm{Im}(\nabla_\Theta \phi(u)) - \frac{\mathrm{Re}(\nabla_\Theta \phi(u))}{u} \right) \sin(uk)}{1+u^2} du. \tag{3.9}$$

Therefore, to calculate the gradient of the objective function in (3.4) with respect to the parameter vector $\boldsymbol{\Theta}$, we need the gradient of the characteristic function ϕ, which is presented below through Theorem 3.1 and its corollary.

Theorem 3.1. *Define*

$$\tilde{\phi}(\omega; u) \equiv e^{iux} \frac{\Gamma(\beta - \alpha)}{\Gamma(\beta)} \zeta^\alpha M\left(\alpha; \beta; -\zeta\right),$$

where α, β and z are univariate functions of ω, with

$$\alpha' \equiv \frac{d\alpha}{d\omega}, \quad \beta' \equiv \frac{d\beta}{d\omega}, \quad \zeta' \equiv \frac{d\zeta}{d\omega}.$$

Then we have that

$$\Omega_\omega \equiv \frac{\partial \tilde{\phi}}{\partial \omega} = \tilde{\phi}(\omega; u) \Big[(\beta' - \alpha')\psi(\beta - \alpha) - \beta' \psi(\beta) + \left(\alpha' \ln(\zeta) + \alpha \frac{\zeta'}{-\zeta} \right) + \frac{\beta G(\alpha; \beta; -\zeta) - \alpha \zeta' M(\alpha+1; \beta+1; -\zeta)}{\beta M(\alpha; \beta; -\zeta)} \Big], \tag{3.10}$$

where ψ is the digamma function, and

$$G(\alpha;\beta;-\zeta) \equiv \sum_{n=1}^{\infty}\left(\alpha' H_n^{\alpha} - \beta' H_n^{\beta}\right)\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!},$$

$$H_n^{\alpha} \equiv \sum_{k=0}^{n-1}\frac{1}{\alpha+k}, \quad H_n^{\beta} \equiv \sum_{k=0}^{n-1}\frac{1}{\beta+k}.$$

Proof. First, recall that the gamma function $\Gamma(x)$ and the digamma function

$$\psi(x) \equiv \frac{\Gamma'(x)}{\Gamma(x)},$$

have the following properties, respectively:

$$(x)_n = \frac{\Gamma(x+n)}{\Gamma(x)},$$

and

$$\psi(x+n) = \psi(x) + \sum_{k=0}^{n-1}\frac{1}{x+k}.$$

With this in mind, we see that

$$\begin{aligned}\frac{\partial}{\partial\omega}\left[\frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\right] &= \frac{(\beta'-\alpha')\psi(\beta-\alpha)\Gamma(\beta-\alpha)\Gamma(\beta) - \beta'\psi(\beta)\Gamma(\beta-\alpha)\Gamma(\beta)}{\Gamma(\beta)^2}\\ &= \frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\left((\beta'-\alpha')\psi(\beta-\alpha) - \beta'\psi(\beta)\right),\end{aligned}$$

and

$$\begin{aligned}&\frac{\partial}{\partial\omega}\left[\sum_{n=0}^{\infty}\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!}\right]\\ =&\frac{\partial}{\partial\omega}\left[\sum_{n=0}^{\infty}\frac{\Gamma(\alpha+n)}{\Gamma(\beta+n)}\frac{\Gamma(\beta)}{\Gamma(\alpha)}\frac{(-\zeta)^n}{n!}\right]\\ =&(\beta'\psi(\beta)-\alpha'\psi(\alpha))\left(M(\alpha,\beta;-\zeta)-1\right)\\ &+\sum_{n=1}^{\infty}\left(\alpha'(\psi(\alpha)+H_n^{\alpha}) - \beta'(\psi(\beta)+H_n^{\beta})\right)\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!} - \sum_{n=0}^{\infty}n\zeta'\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^{n-1}}{n!}\\ =&\sum_{n=1}^{\infty}\left(\alpha' H_n^{\alpha} - \beta' H_n^{\beta}\right)\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!} - \frac{\alpha}{\beta}\zeta' M(\alpha+1,\beta+1,-\zeta).\end{aligned}$$

Furthermore, we have that

$$\frac{\partial\zeta^{\alpha}}{\partial\omega} = \zeta^{\alpha}\left(\alpha'\ln(\zeta) + \frac{\alpha\zeta'}{\zeta}\right).$$

With these results, we have that the derivative of $\tilde{\phi}$ is given by

$$\begin{aligned}
&\frac{\partial \tilde{\phi}}{\partial \omega} \\
&= \frac{\partial}{\partial \omega}\left[e^{iux} \frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)} \zeta^{\alpha} M\left(\alpha;\beta;-\zeta\right)\right] \\
&= e^{iux}\Bigg(\frac{\partial}{\partial \omega}\left[\frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\right]\zeta^{\alpha} M(\alpha,\beta;-\zeta) + \frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\frac{\partial \zeta^{\alpha}}{\partial \omega} M(\alpha,\beta;-\zeta) \\
&\quad + \frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\zeta^{\alpha}\frac{\partial}{\partial \omega}\left[M(\alpha,\beta;-\zeta)\right]\Bigg) \\
&= \left((\beta'-\alpha')\psi(\beta-\alpha) - \beta'\psi(\beta)\right)\tilde{\phi} + \left(\alpha'\ln(\zeta) + \frac{\alpha\zeta'}{\zeta}\right)\tilde{\phi} \\
&\quad + \left(\sum_{n=1}^{\infty}\left(\alpha' H_n^{\alpha} - \beta' H_n^{\beta}\right)\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!} - \frac{\alpha}{\beta}\zeta' M(\alpha+1,\beta+1,-\zeta)\right)\frac{\tilde{\phi}}{M(\alpha,\beta;-\zeta)} \\
&= \tilde{\phi}\Bigg(\left((\beta'-\alpha')\psi(\beta-\alpha) - \beta'\psi(\beta)\right) + \left(\alpha'\ln(\zeta) + \frac{\alpha\zeta'}{\zeta}\right) \\
&\qquad\qquad + \frac{\beta G(\alpha,\beta;-\zeta) - \alpha\zeta' M(\alpha+1,\beta+1,-\zeta)}{\beta M(\alpha,\beta;-\zeta)}\Bigg).
\end{aligned}$$

□

Note that if α and β are constant with respect to ω, as is the case when we differentiate the characteristic function with respect to either η or ν, we end up with a significantly simpler form.

Corollary 3.2. *The gradient of the characteristic function $\phi(\boldsymbol{\Theta};u)$ with respect to $\boldsymbol{\Theta} = [\kappa,\eta,\nu,\sigma,\rho]$ is given by*

$$\nabla_{\Theta}\phi = [\Omega_{\kappa},\Omega_{\eta},\Omega_{\nu},\Omega_{\sigma},\Omega_{\rho}],$$

where $\Omega_{\kappa},\ldots,\Omega_{\rho}$ are defined in theorem 3.1, and the intermediate derivatives of α,β and ζ with respect to the parameters are given by

$$\frac{\partial \alpha}{\partial \sigma} = \frac{2\kappa}{\sigma^3} - \frac{i\rho u}{\sigma^2} + \frac{\left(\frac{-2\kappa}{\sigma^3} + \frac{i\rho u}{\sigma^2}\right)\left(\frac{1}{2} - \frac{p}{\sigma^2}\right) - \frac{2q}{\sigma^3}}{\sqrt{\left(\frac{1}{2} - \frac{p}{\sigma^2}\right)^2 + \frac{2q}{\sigma^2}}},$$

$$\frac{\partial \alpha}{\partial \kappa} = -\frac{1}{\sigma^2} + \frac{\frac{1}{2\sigma^2} - \frac{p}{\sigma^4}}{\sqrt{\left(\frac{1}{2} - \frac{p}{\sigma^2}\right)^2 + \frac{2q}{\sigma^2}}},$$

$$\frac{\partial \alpha}{\partial \rho} = \frac{iu}{\sigma} + \frac{\frac{-iu}{2\sigma} + \frac{iup}{\sigma^3}}{\sqrt{\left(\frac{1}{2} - \frac{p}{\sigma^2}\right)^2 + \frac{2q}{\sigma^2}}},$$

$$\frac{\partial \beta}{\partial \sigma} = 2\left(\frac{\partial \alpha}{\partial \sigma} - \frac{2\kappa}{\sigma^3} + \frac{i\rho u}{\sigma^2}\right),$$

$$\frac{\partial \beta}{\partial \kappa} = 2\left(\frac{\partial \alpha}{\partial \kappa} + \frac{1}{\sigma^2}\right),$$

$$\frac{\partial \beta}{\partial \rho} = 2\left(\frac{\partial \alpha}{\partial \rho} - \frac{iu}{\sigma}\right),$$

$$\frac{\partial \zeta}{\partial \sigma} = -\frac{2\zeta}{\sigma},$$

$$\frac{\partial \zeta}{\partial \nu} = -\frac{\zeta}{\nu},$$

$$\frac{\partial \zeta}{\partial \kappa} = -\left(\eta(T-t) - \frac{1}{\kappa}\right)\zeta - \frac{(T-t)\nu\sigma^2}{2\kappa}\zeta^2,$$

$$\frac{\partial \zeta}{\partial \eta} = -\left(\kappa(T-t) - \frac{1}{\eta}\right)\zeta - \frac{(T-t)\nu\sigma^2}{2\eta}\zeta^2,$$

with all other partial derivatives equal to zero.

3.2.2 Optimal Gradient Computation

In computational terms, the most naive approach to solving (3.4) is to apply an optimization algorithm with finite-difference gradient approximations, and where $\pi(\boldsymbol{\Theta}; (K,T)_i)$ is evaluated individually for each option i without any computational caching. In contrast, the Fast Fourier Transform method of Carr and Madan [1999] computes the option price for an entire grid of strikes for a given maturity simultaneously. The integrand of the original inversion formulation proposed in Heston [1993] precludes the use of the FFT since a singularity is present at $u = 0$. To address this, Carr and Madan [1999] proposed the damped formulation

$$\psi_T(u) = \frac{e^{-r(T-t)}\phi(u - i(\gamma+1))}{\gamma^2 + \gamma - u^2 + i(2\gamma+1)u},$$

where T is the maturity of the option set, and γ is a damping coefficient, which needs to be chosen by hand.[8] Given $\psi_T(\cdot)$, the option price is given by

$$\pi(\mathbf{\Theta}; K_u, T) \approx \frac{e^{-\gamma k_u}}{\pi} \sum_{j=1}^{N} \delta_j e^{-i\lambda\mu(j-1)(u-1)} e^{ibv_j} \psi_T(v_j)\mu, \tag{3.11}$$

where $k_u = -b + \lambda(u-1)$, $v_j = \mu(j-1)$, $b = \frac{1}{2}N\lambda$, $K_u = e^{k_u}$, and $\delta_j = 0.5$ for $j = 1, N$, and $\delta_j = 1$ otherwise.

In words, λ is the distance between consecutive log strike points, μ is the distance between consecutive integration nodes, and $-b$ and b are the lower and upper limits on the log strike range, respectively. Using the FFT to compute the inversion, we set $\lambda\mu = \frac{2\pi}{N}$. This implies a tradeoff between the resolution of the integration grid, and the resolution of the strike price grid.

The FFT method has proven useful for a variety of interesting applications in quantitative finance (see e.g. Albanese et al. [2004]; Chen and Chen [2014]; Fusai et al. [2016]). However, a major drawback of the method in our setting is that a very large number of integration nodes are required to obtain a satisfactory level of accuracy, since the integration nodes need to be equally spaced and that prohibits the use of genuinely efficient integration schemes. Furthermore, we end up with a far greater number of option prices for each maturity than for which we could hope to have any use for most calibration purposes. This is because strike prices far in or out of the money tend to be very illiquid. In other words, most of the option prices we obtain from the FFT are likely matched to noise in the market data which we would prefer to keep out of the calibration.

A different method of tackling the inversion problem was introduced by Fang and Oosterlee [2008]. It is generally referred to as the Fourier-cosine series expansion (COS) method. Here, the inversion integral is replaced by its cosine series expansion, with the series coefficients extracted directly from the integrand. Using this method, we have that

$$\pi(\mathbf{\Theta}; K, T) \approx e^{-r(T-t)} \sum_{j=1}^{N} \delta_j \mathrm{Re}\left\{\phi\left(\frac{j\pi}{b-a}\right) e^{-ij\pi\frac{a}{b-a}}\right\} V_j, \tag{3.12}$$

where a and b are the lower and upper endpoints of the truncated interval of support for the density function, and V_j depends on the payoff of the derivative. For plain vanilla options, V_j can be obtained in closed form.

8. The choice of this parameter has a significant impact on the accuracy of the option prices produced by the model. Furthermore, the optimal value of this parameter depends on the option maturity, so ideally a different damping parameter value is required for each maturity present in the set of benchmark options when calibrating the model.

Define

$$\chi_j(c,d) \equiv \frac{1}{1+\left(\frac{j\pi}{b-a}\right)^2}\left[\cos\left(j\pi\frac{d-a}{b-a}\right)e^d - \cos\left(j\pi\frac{c-a}{b-a}\right)e^c + \frac{j\pi}{b-a}\sin\left(j\pi\frac{d-a}{b-a}\right)e^d - \frac{j\pi}{b-a}\sin\left(j\pi\frac{c-a}{b-a}\right)e^c\right], \quad (3.13)$$

and

$$\psi_j(c,d) \equiv \begin{cases}\left[\sin\left(j\pi\frac{d-a}{b-a}\right) - \sin\left(j\pi\frac{c-a}{b-a}\right)\right]\frac{b-a}{j\pi} & \text{if } j \neq 0,\\ (d-c), & \text{if } j = 0.\end{cases}$$

The coefficients for a call and a put option for a given strike are given, respectively, by

$$V_j^{\text{call}} = \frac{2}{b-a}K(\chi_j(0,b) - \psi_j(0,b)),$$

and

$$V_j^{\text{put}} = \frac{2}{b-a}K(-\chi_j(a,0) + \psi_j(a,0)).$$

The COS method has been reported to be around 20-40 times faster (Fang and Oosterlee [2008]) than the FFT method[9]. Its efficiency depends in part on the maturity, with longer maturities requiring a lower N than shorter maturities. Another factor is the choice of the truncation, i.e., the interval $[a,b]$. In Fang and Oosterlee [2008], the following is proposed for the Heston model for maturities T between 0.1 and 10.0:

$$[a,b] \equiv \left[c_1 - L\sqrt{c_2+\sqrt{c_4}}, c_1 + L\sqrt{c_2+\sqrt{c_4}}\right],$$

where c_n is the n-th cumulant of $\ln(S_T/K)$, and L is a control parameter which in Fang and Oosterlee [2008] is set equal to 10.

A somewhat simpler method, which is of similar computational efficiency as the COS method, is proposed in Kilin [2011]. By noting that the characteristic function ϕ does not depend on the strike price K, we see that once we have computed the inversion of ϕ for a given option i we can reuse the computed values of ϕ to quickly compute (3.8) for the remainder of the options that have the same maturity T as i but different strike prices. In addition, we can couple this approach with an efficient integration scheme, like Gaussian quadrature.

9. We find this difference to be smaller with a more streamlined implementation of the FFT. We would like to thank an anonymous referee for providing us with the ideas behind making the FFT faster.

To be more specific, a direct Fourier inversion of the characteristic function in (3.8) means calculating a discrete approximation to the integral in (3.8), which we can write as

$$\sum_{j=1}^{J} b_j \frac{\left(\mathrm{Re}(\phi(u_j)) + \frac{\mathrm{Im}(\phi(u_j))}{u_j}\right)\cos(u_j k) + \left(\mathrm{Im}(\phi(u_j)) - \frac{\mathrm{Re}(\phi(u_j))}{u_j}\right)\sin(u_j k)}{1+u_j^2}, \tag{3.14}$$

where $b_1,\ldots,b_J$ are the integration weights, and are determined by the integration scheme, along with the discretization $u_1,\ldots,u_J$ of the integrand domain.

In a similar fashion, the numerical inversion of the analytic gradient of the characteristic function can be written as

$$\sum_{j=1}^{J} b_j \frac{\left(\mathrm{Re}(\nabla_\Theta\phi(u_j)) + \frac{\mathrm{Im}(\nabla_\Theta\phi(u_j))}{u_j}\right)\cos(u_j k) + \left(\mathrm{Im}(\nabla_\Theta\phi(u_j)) - \frac{\mathrm{Re}(\nabla_\Theta\phi(u_j))}{u_j}\right)\sin(u_j k)}{1+u_j^2}. \tag{3.15}$$

We first give an outline of the technique of computing the prices of a set of vanilla options through caching the characteristic function calculations for each maturity without the use of an analytic gradient, as described in Kilin [2011]. We refer to this method hereafter as strike caching (SC). Let the set of different maturities present in the set of options be denoted by $\boldsymbol{T} = \{T_s\}_{s=1}^{S}$, and for a given maturity T_s in this set, let $\boldsymbol{K}^s = \left\{K_r^s\right\}_{r=1}^{R_s}$ be the set of different strike prices present in the set of benchmark options with maturity T_s.

The method works for a given maturity T_s by calculating and caching the grid $\boldsymbol{A} = \left[A_1,\ldots,A_J\right]$ where $A_j = \phi(u_j)$ for $j = 1,\ldots,J$. This grid is then used for the computation of (3.14) for each strike K_r^s, which reduces the number of calls to ϕ by a factor equal to the number of strikes for each maturity. To make this work, we need to use the same discretization grid for every option with maturity T_s.

If we use adaptive quadrature to carry out the inversion, the discretization grid $\boldsymbol{U} = [u_1,\ldots,u_J]$, and the integration weights $\boldsymbol{B} = [b_1,\ldots,b_J]$, can be obtained simply through keeping track of the node-weight pairs generated by the quadrature from the calculation of the first option of each maturity. Alternatively, we can create the discretization in advance by standard interpolation methods. The application of Algorithm 1 to a finite difference gradient based minimization of (3.4) is straightforward. Let $\boldsymbol{e} = [1,\ldots,1]$ be a vector of the same dimension as $\boldsymbol{\Theta}$, and let h be a sufficiently small, positive number. For a central finite difference we simply evaluate $SC(\cdot)$ for the parameter vectors $\boldsymbol{\Theta} + h\boldsymbol{e}$ and $\boldsymbol{\Theta} - h\boldsymbol{e}$, and then divide the difference of the two return values with $2h$, to obtain the gradient of (3.4) at $\boldsymbol{\Theta}$.

Algorithm 1 Price Computation with Strike Caching

procedure SC($\boldsymbol{\Theta}, \{(K,T)_i\}_{i=1}^{I}$)
  for $s = 1,\ldots,S$ **do**
    Compute $\boldsymbol{U}$ and $\boldsymbol{B}$
    for $j = 1,\ldots,J$ **do**
      $A_j \leftarrow \phi_{T_s}(\boldsymbol{\Theta}; u_j)$
    end for
    for $r = 1,\ldots,R^s$ **do**
      Compute (3.14) with $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{U}$
      $\pi_i \leftarrow \pi(\boldsymbol{\Theta}; K_r^s, T_s)$ using (3.8)
    end for
  end for
  return $\sum_{i=1}^{I} w_i \left(\pi_i - \pi_i^{\text{Market}}\right)^2$
end procedure

The strike caching approach, of course, works just as well for efficiently computing the analytic gradient of the characteristic function. To gain further performance improvements, however, we extend the caching concept, which so far has been limited to the strike price dimension, to include the maturity dimension as well.

While the characteristic function ϕ does not depend on the option strike price, which makes the strike caching technique straightforward, it does indeed depend on the option maturity. However, if we take a closer look at (3.7) we see that only ζ depends on the maturity. So, calculating $\phi(\boldsymbol{\Theta}; u_j)$ simultaneously for the entire set $\boldsymbol{T}$ while keeping u_j fixed enables us to significantly reduce the overall computational load, since α and β remain the same for each $T_s \in \boldsymbol{T}$.

Coupling this approach with the computation of the analytic gradient we presented in the previous section yields an even greater speedup of the calibration procedure. As is apparent from Theorem 3.1 and its corollary, the partial derivatives of the gradient of the characteristic function of the 3/2 model tend to include common factors. This means we can use the results from the computation of one partial derivative to help us calculate the remaining ones.

The application of the strike-maturity caching technique to the calibration problem with the analytic gradient of ϕ will be referred to hereafter as the gradient-maturity-strike caching (GMSC) method.

Let $\Xi_{\alpha,\beta}$ denote the Jacobian of (α,β) with respect to the set of calibration parameters, i.e.,

$$\Xi_{\alpha,\beta} = \begin{bmatrix} \dfrac{\partial \alpha}{\partial \omega_1} & \cdots & \dfrac{\partial \alpha}{\partial \omega_5} \\ \dfrac{\partial \beta}{\partial \omega_1} & \cdots & \dfrac{\partial \beta}{\partial \omega_5} \end{bmatrix},$$

where $\omega_i \in \{\kappa, \eta, v, \sigma, \rho\}$. Similarly, let $\Xi_{\boldsymbol{\zeta}}$ denote the Jacobian of $\zeta_1, \ldots, \zeta_S$ with respect to the model parameters, where ζ_s, for $s = 1, \ldots, S$, is calculated using maturity $T_s \in \boldsymbol{T}$, i.e.,

$$\Xi_{\boldsymbol{\zeta}} = \begin{bmatrix} \dfrac{\partial \zeta_1}{\partial \omega_1} & \cdots & \dfrac{\partial \zeta_1}{\partial \omega_5} \\ \vdots & \ddots & \\ \dfrac{\partial \zeta_S}{\partial \omega_1} & \cdots & \dfrac{\partial \zeta_S}{\partial \omega_5} \end{bmatrix}.$$

Then the GMSC method is given by Algorithm 2.

The order in which the calculations are done in Algorithm 2 is designed to take advantage of cached intermediate results. In line 2 we compute the integration nodes for the transform inversion, and in line 3 we compute $\boldsymbol{\zeta}$ and $\Xi_{\boldsymbol{\zeta}}$, which are independent of the integration nodes. In line 5 we compute α, β and $\Xi_{\boldsymbol{\alpha},\boldsymbol{\beta}}$, that are independent of the option maturity. In lines 6-10 we calculate the grid that consists of the characteristic function values for integration nodes $u_1, \ldots, u_J$, as well as the partial derivatives of the characteristic function in those nodes. The characteristic function evaluation can be reused for the gradient computation, and is returned along with the gradient in our actual implementation, which speeds up the optimization procedure as well. Once we have computed the characteristic function and its gradient for the grid values $\boldsymbol{U}$, we use these stored values to calculate the integral required in (3.8) and (3.9), for the different strikes and maturities present in the set of benchmark options, as written in lines 13-22. Finally, the algorithm returns the vector of partial derivatives of the objective function with respect to the model parameters and spot volatility.

3.3 Regularization with Risk-Neutral MCMC Estimation

The objective function that corresponds to the nonlinear least squares formulation given by (3.4) turns out to exhibit numerically signs of being both non-convex and ill-conditioned, as we discuss in greater detail in section 3.4. One approach to addressing this issue is to regularize (3.4) with a penalty

Algorithm 2 Gradient-Maturity-Strike Caching

procedure GMSC($\boldsymbol{\Theta}, \{(K,T)_i\}_{i=1}^{I}$)
  `Compute` $\boldsymbol{U}$ `and` $\boldsymbol{B}$
  `Compute` $\boldsymbol{\zeta}$ `and` $\Xi_{\boldsymbol{\zeta}}$
  for $j = 1,\dots,J$ **do**
    Compute $\alpha(u_j), \beta(u_j)$, and $\Xi_{\alpha,\beta}(u_j)$
    for $s = 1,\dots,S$ **do**
      $A_{s,j} \leftarrow \phi_{T_s}(u_j)$
      for $g = 1,\dots,5$ **do**
        $G_{s,j,g} \leftarrow \dfrac{\partial \phi_{T_s}(u_j)}{\partial \omega_g}$
      end for
    end for
  end for
  for $s = 1,\dots,S$ **do**
    for $r = 1,\dots,R$ **do**
      `Compute` (3.14) `with` $\phi_{T_s}(u_j) = A_{j,s}$ `for` $j = 1,\dots,J$
      $\pi_i \leftarrow \pi(\boldsymbol{\Theta}; K_r, T_s)$ `using` (3.8)
      for $g = 1,\dots,5$ **do**
        `Compute` (3.15) `with` $\dfrac{\partial \phi_{T_s}(u_j)}{\partial \omega_g} = G_{j,s,g}$ `for` $j = 1,\dots,J$
        $D_{i,g} \leftarrow \dfrac{\partial \pi(\boldsymbol{\Theta}; K_r, T_s)}{\partial \omega_g}$ `using` (3.9)
      end for
    end for
  end for
  return $\left[\sum_{i=1}^{I} w_i \left(\pi_i - \pi_i^{\text{Market}}\right) D_{i,1}, \dots, \sum_{i=1}^{I} w_i \left(\pi_i - \pi_i^{\text{Market}}\right) D_{i,5}\right]$
end procedure

function. In other words, we change the problem from (3.4) to the regularized version

$$\mathbf{\Theta}^* = \underset{\mathbf{\Theta}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{I} w_i f_i(\mathbf{\Theta})^2 + \chi H(\check{\mathbf{\Theta}} - \hat{\mathbf{\Theta}}) \right\}. \tag{3.16}$$

Here, $\hat{\mathbf{\Theta}} = (\hat{\omega}_1, \ldots, \hat{\omega}_r)$, with $r \leq |\mathbf{\Theta}|$, is a regularization point for the parameter vector $\check{\mathbf{\Theta}} = (\omega_1, \ldots, \omega_r)$, with $\omega_i \in \mathbf{\Theta}$ for $i = 1, \ldots, r$. Furthermore, H is a function that penalizes solutions to (3.16) in proportion to how far $\check{\mathbf{\Theta}}$ is from the regularization point, $\hat{\mathbf{\Theta}}$. Finally, χ is the regularization coefficient which determines how severely the penalty impacts the calibration. When χ is set large enough, the objective function becomes convex in $\check{\mathbf{\Theta}}$. If it is too high, however, potentially good solutions outside the immediate vicinity of the regularization point are ignored. The question then arises which χ, $\hat{\mathbf{\Theta}}$, and H we should choose.

One obvious choice for $\hat{\mathbf{\Theta}}$ is a historical estimate of the proper model parameters (Cont and Tankov [2004]). In other words, we estimate $\{\kappa, \eta, \sigma, \rho\}$ using the time series of prices of the underlying realized over time. On a conceptual level, the reason why we want historical data in the calibration is to make the calibration process more robust against the noise inherent in option prices. This noise comes from sources such as the bid-ask spreads, "easy number pricing" (i.e., the market prices of options tend to be given in convenient numbers at the expense of accuracy), and asynchronous pricing in end-of-day data. Option prices are therefore only an imperfect proxy for the market pricing measure.

When the estimation of the model parameters is given by (3.4), the parameter values we obtain correspond to the risk neutral measure, since the benchmark options are themselves priced under that measure. Estimation of the parameters under the objective measure, however, involves a considerably different formulation of the problem, and requires a study of the historical realization of the price of the underlying. This is relatively straightforward for discrete-time models, such as generalized autoregressive conditional heteroskedasticity (GARCH) and its variants, where maximum likelihood can be applied directly to the model. In contrast, the application of standard econometric techniques to continuous-time models are complicated by the fact that the density is usually not known for the model, so maximum likelihood is not directly applicable. In addition, the instantaneous volatility is assumed to be unobserved in the 3/2 model, which further complicates the estimation.

Several methods have been developed in recent years to tackle this problem for continuous-time stochastic volatility models, such as 'implied state' generalized method of moments (GMM) (Pan [2002]), and simulated maximum

likelihood (Brandt and Santa-Clara [2002]; Pedersen [1995]). The approach we have chosen is Bayesian inference through the Markov chain Monte Carlo method, see Johannes and Polson [2009] for an introduction. Let $\{S_t\}_{t=0}^T$ be a realized time series of prices for an underlying asset, whose dynamics are given by (3.2), and let $Y_t = \ln S_t - \ln S_{t-1}$, and $\boldsymbol{Y} = \{Y_t\}_{t=1}^T$. Assume further, for the time being, that we do not observe the corresponding realization of the time series of instantaneous volatility, $\boldsymbol{V} = \{V_t\}_{t=1}^T$. Lastly, let $\tilde{\boldsymbol{\Theta}} = (\tilde{\omega}_1, \ldots, \tilde{\omega}_r)$ be the parameters we want to estimate using the historical data, interpreted in the conventional Bayesian sense (i.e., as stochastic variables). Estimating $\tilde{\boldsymbol{\Theta}}$ using the MCMC approach involves repeatedly drawing samples from the distribution given by the posterior distribution $p(\tilde{\boldsymbol{\Theta}}, \boldsymbol{V}|\boldsymbol{Y})$. More specifically, assume we have drawn n samples using the posterior distribution, then the $(n+1)$-th sample for $\tilde{\boldsymbol{\Theta}}$ is drawn in the following way,

$$
\begin{aligned}
&\text{1. Draw } \tilde{\Theta}_1^{(n+1)} \sim p\left(\tilde{\Theta}_1 | \tilde{\Theta}_2^{(n)}, \tilde{\Theta}_3^{(n)}, \ldots, \tilde{\Theta}_r^{(n)}, \boldsymbol{V}^{(n)}\right),\\
&\text{2. Draw } \tilde{\Theta}_2^{(n+1)} \sim p\left(\tilde{\Theta}_2 | \tilde{\Theta}_1^{(n+1)}, \tilde{\Theta}_3^{(n)}, \ldots, \tilde{\Theta}_r^{(n)}, \boldsymbol{V}^{(n)}\right),\\
&\vdots\\
&\text{r. Draw } \tilde{\Theta}_r^{(n+1)} \sim p\left(\tilde{\Theta}_r | \tilde{\Theta}_1^{(n+1)}, \tilde{\Theta}_2^{(n+1)}, \ldots, \tilde{\Theta}_{r-1}^{(n+1)}, \boldsymbol{V}^{(n)}\right).
\end{aligned}
$$

With $\boldsymbol{V}$ unobserved we would then, in a similar fashion, also have to repeatedly sample the spot volatilities, using $p\left(V_t | \tilde{\boldsymbol{\Theta}}^{(n)}, V_1^{(n)}, \ldots, V_T^{(n)}\right)$ to generate $V_t^{(n+1)}$, for each $t = 1, \ldots, T$. As is (currently) the case for most stochastic volatility models, we do not have a tractable form for $p(V_t|\cdots)$, so an approximate procedure, such as the Metropolis-Hastings algorithm, is required to sample it. This is by far the most computationally challenging aspect of the MCMC estimation.

Note that the procedure outlined above only gives us estimates for the parameters of the model under the objective measure. This can be useful, for instance, if we are concerned with how well the model fits the realized price history of the underlying. However, (3.4) is formulated for option prices, which means the parameters we are after are under the risk neutral measure. This is not an issue for σ and ρ which, by Girsanov's theorem, stay the same under each measure, whereas η and κ change. To obtain the risk neutral version of η and κ we need to include option prices in the historical estimation as well to estimate the variance risk premium.

To make this point clear, assume that the market price of volatility risk is given by $\nu V^{1/2}$, i.e., it is linear in volatility[10], and let W_t^p and W_t^q be Brownian motions under $\mathbb{P}$ and $\mathbb{Q}$, respectively. To further clarify our discussion,

10. This is a standard simplifying assumption in both the Heston model and the 3/2 model, as it

denote with κ^p and κ^q the rate of mean reversion under $\mathbb{P}$ and $\mathbb{Q}$, respectively, and η^p and η^q the long run volatility under $\mathbb{P}$ and $\mathbb{Q}$, respectively (as mentioned above σ and ρ stay the same).

From Girsanov's theorem, we have that under a change of measure from $\mathbb{P}$ to $\mathbb{Q}$ the volatility process

$$dV_t = \kappa^p(\eta^p V_t - V_t^2)\,dt + \sigma V_t^{3/2} dW_t^p,$$

becomes

$$\begin{aligned} dV_t &= \kappa^p(\eta^p V_t - V_t^2)\,dt + \sigma V_t^{3/2} dW_t^q - \sigma\nu V_t^2\,dt \\ &= \left(\kappa^p + \sigma\nu\right)\left(\frac{\kappa^p\eta^p}{\kappa^p + \sigma\nu} V_t - V_t^2\right) dt + \sigma V_t^{3/2} dW_t^q \\ &= \kappa^q(\eta^q V_t - V_t^2)\,dt + \sigma V_t^{3/2} dW_t^q. \end{aligned}$$

As we can see from this derivation,

$$\kappa^q = \kappa^p + \sigma\nu, \tag{3.17}$$

and

$$\eta^q = \frac{\kappa^p\eta^p}{\kappa^p + \sigma\nu}. \tag{3.18}$$

so ν affects the rate of mean reversion and long run volatility when we change from $\mathbb{P}$ to $\mathbb{Q}$, which means that if we want historical regularization values for the full set of the model parameters under $\mathbb{Q}$ then we need to estimate ν.

Estimation of the variance risk premium is in itself an important part of the study of incomplete markets, for which MCMC estimation is also suitable. There is, however, the issue that incorporating option pricing into the MCMC estimation is very computationally challenging. The intuition we present here is that because we are only interested in the historical estimates of the parameters of the 3/2 model under the risk neutral measure, we can simply cut out the intermediate step of estimating the parameters and volatility process under the objective measure altogether by using historical implied volatility as a proxy for the volatility process under $\mathbb{Q}$. In other words, instead of filtering V_t under $\mathbb{P}$ and jointly estimating ν by incorporating past option prices and using formulas (3.17) and (3.18) to find the historical estimates of κ and η under $\mathbb{Q}$, we simply take the implied volatility of short dated options as a proxy for V_t and jump straight into the estimation of the 3/2 model parameters under $\mathbb{Q}$.

implies these models are the same under the objective measure and risk neutral measure, as well as having economic justification, see Heston [1993].

While this idea can be implemented in several different ways, we use CBOEs volatility index VIX for our numerical tests. The VIX is the square root of the risk-neutral expectation of the S&P500 variance over the following 30 calendar days, as determined by the market.[11]

With the proxy for $\boldsymbol{V}$, we now only need to sample $p(\tilde{\boldsymbol{\Theta}}|\boldsymbol{Y},\boldsymbol{V})$. And, as we show below, we can characterize the posterior distribution for each parameter in $\tilde{\boldsymbol{\Theta}}$ in closed form, which makes the MCMC estimation process of the historical data extremely fast.

We begin by discretizing (3.2) to obtain

$$\begin{aligned} Y_t &= \left(r - \frac{1}{2}V_{t-1}\right)\Delta t + \sqrt{\Delta t V_{t-1}}\epsilon_t^s, \\ V_t &= V_{t-1} + V_{t-1}\kappa(\eta - V_{t-1})\Delta t + V_{t-1}^{3/2}\sqrt{\Delta t}\epsilon_t^v, \end{aligned} \tag{3.19}$$

where $\epsilon_t^s \sim N(0,1)$ and $\epsilon_t^v \sim N(0,\sigma)$. Note that as we are treating the process (3.19) under the risk neutral measure, the drift of Y is given by the risk free rate, r. The equations above can be rewritten in terms of the errors, to give

$$\begin{aligned} \epsilon_t^s &= \frac{Y_t - r\Delta t + \frac{1}{2}V_{t-1}\Delta t}{\sqrt{V_{t-1}\Delta t}}, \\ \epsilon_t^v &= \frac{V_t - V_{t-1} - V_{t-1}\kappa(\eta - V_{t-1})\Delta t}{V_{t-1}^{3/2}\Delta t}. \end{aligned}$$

Following the setup of Li et al. [2006] for the MCMC analysis of affine jump-diffusion, we change variables and set $\psi = \rho\sigma$, and $\Omega = \sigma^2(1-\rho^2)$. This gives us

$$(\epsilon_t^s, \epsilon_t^v) \sim N\left((0,0), \begin{bmatrix} 1 & \psi \\ \psi & \psi^2 + \Omega \end{bmatrix}\right).$$

We now require the posterior distributions, $p(\boldsymbol{\Theta}|\boldsymbol{Y},\boldsymbol{V})$, which we use to sample the parameter space. By Bayes' theorem,

$$p(\tilde{\omega}_i|\tilde{\boldsymbol{\Theta}}_{-i},\boldsymbol{V},\boldsymbol{Y}) \propto p(\boldsymbol{Y},\boldsymbol{V}|\tilde{\omega}_i,\tilde{\boldsymbol{\Theta}}_{-i})p(\tilde{\omega}_i),$$

for $i = 1,\ldots,5$, where $\tilde{\boldsymbol{\Theta}}_{-i} \equiv \tilde{\boldsymbol{\Theta}} \setminus \{\tilde{\omega}_i\}$. The joint likelihood function for $(\boldsymbol{Y},\boldsymbol{V})$ is given by

$$p(\boldsymbol{Y},\boldsymbol{V}|\tilde{\boldsymbol{\Theta}}) = \frac{1}{\Omega^{T/2}}\left(\prod_{t=1}^{T}\frac{1}{V_{t-1}^2\Delta T}\right)\exp\left(-\frac{1}{2\Omega}\sum_{t=1}^{T}\left((\Omega+\psi^2)(\epsilon_t^s)^2 - 2\psi\epsilon_t^s\epsilon_t^v + (\epsilon_t^v)^2\right)\right).$$

11. In theory, we could choose a volatility index with any time horizon, and deduce the instantaneous volatility by solving $(\text{VIX}/100)^2 = \mathbb{E}^{\mathbb{Q}}\left[\int_t^T v_u du | V_t = v\right]$ for V_t. However, while we know the right hand side in closed form (Carr and Sun [2007]), the expression is complicated enough that we would in fact be better off simply including estimation of $\boldsymbol{V}$ in the MCMC procedure.

Following standard results on the conjugate prior for a normal distribution with unknown mean and variance (Koch [2007]), the prior we choose for Ω is the inverse gamma distribution, i.e., $\Omega \sim \mathcal{IG}(\alpha_0, \beta_0)$, and for the prior for ψ conditioned on Ω we choose the normal distribution, i.e., $\psi|_\Omega \sim N(\psi_0, p_0\Omega)$.

With these priors, the posterior distribution for Ω is given by $\mathcal{IG}(\widehat{\alpha}, \widehat{\beta})$, where $\widehat{\alpha} = \frac{T}{2} + \alpha_0$, and

$$\widehat{\beta} = \beta_0 + \frac{1}{2}\sum_{t=1}^{T}(\epsilon_t^v)^2 + \frac{p_0\psi_0^2}{2} - \frac{p_0\psi_0 + \sum_{t=1}^{T}(\epsilon_t^s\epsilon_t^v)^2}{p_0 + \sum_{t=1}^{T}(\epsilon_t^2)^2}.$$

The posterior conditional distribution for ψ given Ω is given by $N(\widehat{\psi}, \widehat{\sigma}_\psi^2)$, where

$$\widehat{\psi} = \frac{p_0 + \sum_{t=1}^{T}\epsilon_t^s\epsilon_t^v}{p_0 + \sum_{t=1}^{T}\left(\epsilon_t^s\right)^2},$$

and

$$\widehat{\sigma}_\psi^2 = \frac{\Omega}{p_0 + \sum_{t=1}^{T}(\epsilon_t^s)^2}.$$

These posteriors turn out to be the same for the Heston model. The remaining estimators, i.e., for η and κ, however, turn out to be different.

The prior we choose for κ is the normal distribution, i.e., $\kappa \sim N(\kappa_0, \sigma_\kappa^2)$. The posterior is then given by $N(\widehat{\kappa}, \widehat{\sigma}_\kappa^2)$, where

$$\widehat{\kappa} = \frac{\widehat{\sigma}_\kappa^2}{\Omega}\sum_{t=1}^{T}\frac{(\eta - V_{t-1})((\eta - V_{t-1})V_{t-1}\Delta t - \psi(Y_t - r\Delta t + \frac{1}{2}V_{t-1}\Delta t) - 1)}{V_{t-1}} + \frac{\widehat{\sigma}_\kappa^2}{\Omega}\sum_{t=1}^{T}\frac{V_t(\eta - V_{t-1})}{V_{t-1}^2} + \frac{\widehat{\sigma}_\kappa^2\kappa_0}{\sigma_\kappa^2},$$

and

$$\widehat{\sigma}_\kappa^2 = \frac{1}{\sum_{t=1}^{T}\frac{(V_{t-1} - \eta)^2\Delta t}{\Omega V_{t-1}} + \frac{1}{\sigma_\kappa^2}}.$$

Likewise, the prior we pick for η is the normal distribution, i.e., $\eta \sim N(\eta_0, \sigma_\eta^2)$, and the posterior is given by $N(\widehat{\eta}, \widehat{\sigma}_\eta^2)$, where

$$\widehat{\eta} = \frac{\widehat{\sigma}_\eta^2}{\Omega}\sum_{t=1}^{T}\frac{\kappa(\kappa V_{t-1}\Delta t - \psi(Y_t - r\Delta t + \frac{1}{2}V_{t-1}\Delta t) - 1)}{V_{t-1}} + \frac{\widehat{\sigma}_\eta^2}{\Omega}\sum_{t=1}^{T}\frac{V_t\kappa}{V_{t-1}^2} + \frac{\widehat{\sigma}_\eta^2\eta_0}{\sigma_\eta^2},$$

and

$$\widehat{\sigma}_\kappa^2 = \frac{1}{\sum_{t=1}^{T}\frac{\kappa^2\Delta t}{\Omega V_{t-1}} + \frac{1}{\sigma_\eta^2}}.$$

These posterior distributions allow us to sample, and therewith estimate, the parameters κ, η, σ and ρ in terms of historical returns and the implied volatility proxy. One advantage of the MCMC estimation procedure is that we naturally get a gauge for the uncertainty over our parameter estimates. More specifically, from the samples we draw we can calculate the sample covariance matrix

$$\boldsymbol{Q} = \begin{bmatrix} q_{1,1} & \cdots & q_{1,4} \\ \vdots & \ddots & \\ q_{4,1} & \cdots & q_{4,4} \end{bmatrix},$$

where

$$q_{i,j} = \frac{1}{N-1} \sum_{n=1}^{N} (\tilde{\omega}_{i,n} - \overline{\omega}_i)(\tilde{\omega}_{j,n} - \overline{\omega}_j).$$

Here, $\tilde{\omega}_{i,n}$ and $\tilde{\omega}_{j,n}$ are the nth realizations of parameter i and j, respectively, and $\overline{\omega}_i$ and $\overline{\omega}_j$ are their respective sample averages.

We can then utilize this information to determine how to penalize deviations across the individual historical estimates, $\tilde{\omega}_1, \ldots, \tilde{\omega}_4$, of the model parameters during the calibration. For this purpose we have chosen the following parsimonious formulation for the penalty function H.

$$H(\check{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}}) \equiv (\check{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}})^{\top} \boldsymbol{Q}^{-1} (\check{\boldsymbol{\Theta}} - \hat{\boldsymbol{\Theta}}).$$

For a linear least squares problem, this form of the penalty function H is often referred to as Tikhonov regularization (Tikhonov et al. [1979]), and has been studied extensively in the literature.

One computational advantage of the MCMC method relevant to our regularization effort is that if we want to incorporate new data into a prior historical estimate, we do not have to repeat the estimation procedure for the entire data set. More specifically, let $\pi^0(\tilde{\boldsymbol{\Theta}})$ be a prior distribution for the model parameters. Given observations $\boldsymbol{Y} = \{Y_t\}_{t=1}^{T}$ and $\boldsymbol{V} = \{V_t\}_{t=1}^{T}$, let

$$\pi^T(\tilde{\boldsymbol{\Theta}}) \equiv p(\tilde{\boldsymbol{\Theta}} | \boldsymbol{V}, \boldsymbol{Y}) \propto \pi^0(\tilde{\boldsymbol{\Theta}}) p(\boldsymbol{V}, \boldsymbol{Y} | \tilde{\boldsymbol{\Theta}})$$

denote the posterior distribution. If we then acquire an additional observation (Y_{T+1}, V_{T+1}), we have that

$$\pi^{T+1}(\tilde{\boldsymbol{\Theta}}) \propto \pi^T(\tilde{\boldsymbol{\Theta}}) p(Y_{T+1}, V_{T+1} | \boldsymbol{V}, \boldsymbol{Y}, \tilde{\boldsymbol{\Theta}}),$$

where $\pi^{T+1}(\tilde{\boldsymbol{\Theta}})$ is the updated posterior distribution. This means that as time progresses, maintaining an up-to-date historical estimate by adding price movements and volatilities as they get realized is reduced to sampling the parameter space for those incoming values.

Lastly, the choice of χ is done on case-by-case basis. No results exist so far, to our knowledge, that deal definitively with the issue. Roughly speaking, we are after a value for χ which is at least high enough to produce a convex objective function, but not so high that the parameter search space becomes degenerate. Our search for χ is discussed in more detail in section 3.4.

3.4 Numerical Results

Our numerical experiments tested the speed of the GMSC formulation proposed in section 3.2 and summarized in Algorithm 2, as well as the pure SC formulation of Kilin [2011] summarized in Algorithm 1, the COS method of Fang and Oosterlee [2008], and the FFT of Carr and Madan [1999], for a fixed level of accuracy. For this part, we used option prices derived from the 3/2 model with randomly selected parameter values. We also tested the effects of regularization as described in section 3.3 and for this we used S&P500 option data in addition to realized historical returns of the S&P500. The calibration code was written in C/C++ with Alglib and GNU GSL using full compiler optimization (/O2,/Ot) with MSVC. The historical estimation code was written in Python with Numpy and Scipy. The tests were run on an Intel Core i5-4310U (2.0GHz and 2.6GHz) with 8 GB of memory.

3.4.1 Calibration Speed Tests

For the first part of our numerical experiments, we timed the gradient computation of (3.4) using (i) the central finite difference gradient scheme with the FFT method, (ii) the central finite difference gradient scheme with the COS method, (iii) the central finite difference gradient scheme with strike caching, and (iv) the analytic formula for the gradient of the characteristic function where gradient values were cached across both the strike and maturity dimension. Each gradient evaluation was done using 100 different options; 10 strikes per maturity, for a total of 10 maturities per evaluation. The list of maturities consisted of the following: 15, 30, 45, 60, 90, 120, 150, 220, 270 and 365 days. The strikes were close to the money, with strike price increments $\Delta K = 5$ for $T \leq 45$; $\Delta K = 10$ for $45 < T \leq 120$; and $\Delta K = 20$ for $120 < T \leq 365$. The parameter values $\boldsymbol{\Theta}$ were randomly generated on the intervals given in Table 3.1.

Once a random value had been generated for $\boldsymbol{\Theta}$, each of the methods in (i)-(iv) was used to calculate the gradient of (3.4), and the CPU time for each was recorded. This procedure was repeated for 400 different parameter realizations. For the FFT method we used 2^{12} integration points with a spacing

of 0.125. For the SC and GMSC methods we used Gauss-Legendre quadrature with 128 points. For the COS method we used the truncation for $[a, b]$ given in section 3.2.2, where the cumulants were calculated using numerical differentiation of the characteristic function. Furthermore, we used a fully vectorized form for the COS method as given in Fang and Oosterlee [2008]. The error tolerance of the inversion component for both the vectorized COS method and the FFT method[12] was set at 10^{-6}. Lastly, the step size for the finite difference gradient approach was set to 64 bit floating point precision.

In addition, we timed the full nonlinear least squares calibration of the 3/2 model with box constraints using each of these Fourier inversion methods, using the same setup for generating the option data and the true parameter values as in the gradient tests, with the same parameter constraints as given in Table 3.1. The parameter values were randomly generated, as well as the starting point of the optimization, and then each of the methods (i)-(iv) used to solve the inverse problem of retrieving the realized parameter values (each method was tested using the same set of realized parameter values and initial guesses to ensure unbiased comparison). This procedure was repeated 100 times. The optimization routine was set to stop when the gradient changed less than 2/3 of 64 bit floating point precision between iterations. The summary statistics for these tests are given in Table 3.3.

	min	max
κ	1.0	50.0
η	0.01	0.90
ν	0.01	0.90
σ	5.0	50.0
ρ	-1.0	0.50

Table 3.1: Variable Constraints. These represent the intervals within which we constrained the optimization procedure for each parameter.

The calibration converged to the true parameter values and terminated within the specified stopping condition in all 100 test cases for each of the methods (i)-(iv). The mean absolute difference between the least squares parameter estimates and the true parameter values is given in Table 3.4.

As can be seen from Tables 3.2 and 3.3, the GMSC clearly outperforms the other methods both in computing the gradient and in the calibration runs.

12. More specifically, the loop calculating the summands of the vectorized inversion of (3.12), and the inversion of (3.11) was terminated as soon as five consecutive summands turned out to be below the error threshold in absolute value.

	FFT	COS	SC	GMSC
mean	2.621	0.6109	0.5514	0.03519
std	1.730	0.8214	0.5410	0.02320
max	12.79	4.331	3.918	0.1047
min	1.549	0.2988	0.3710	0.01500

Table 3.2: Computational times in seconds for the central finite difference gradient approximation calculations for the FFT method, the COS method, the SC method, and for the analytic gradient with the GMSC method.

	FFT	COS	SC	GMSC
mean	168.3	53.46	48.65	4.499
std	101.9	66.17	43.64	4.013
max	821.7	205.9	174.6	12.17
min	9.562	3.588	1.244	0.0658

Table 3.3: Computational times in seconds of the calibration runs for the FFT method, the COS method, the SC method, and the GMSC method.

We notice a slight decrease in the speed gap for the calibration tests, as compared to the pure gradient computations. This is because the implementation of the Levenberg-Marquardt procedure that we used required in some instances objective function evaluations independent of the gradient evaluations. Overall, the full calibration testing procedure is a somewhat noisier method of measurement than the pure gradient testing procedure, since it introduces the issue of finite accuracy of the finite difference gradient, which can have an implementation-specific effect on the convergence of the algorithm. Also, different implementations of a given optimization algorithm can involve varying memory access overhead which is difficult to control for. From Table 3.4 we see that the GMSC produces slightly more accurate estimates on average than the comparison methods. This is not a surprise since the stopping conditions for the optimization are set right at the cutoff for the accuracy of the central finite difference gradient. However, all the different calibration methods exhibit good accuracy.

	FFT	COS	SC	GMSC
$\lvert\kappa^\dagger - \kappa^*\rvert$	4.10×10^{-2}	3.05×10^{-2}	1.61×10^{-2}	5.16×10^{-3}
$\lvert\theta^\dagger - \theta^*\rvert$	4.73×10^{-4}	3.89×10^{-4}	3.66×10^{-4}	1.19×10^{-4}
$\lvert v^\dagger - v^*\rvert$	8.26×10^{-6}	1.50×10^{-5}	1.38×10^{-6}	6.80×10^{-7}
$\lvert\sigma^\dagger - \sigma^*\rvert$	1.20×10^{-3}	3.25×10^{-3}	9.44×10^{-4}	1.16×10^{-4}
$\lvert\rho^\dagger - \rho^*\rvert$	4.99×10^{-4}	6.05×10^{-4}	2.12×10^{-4}	8.63×10^{-5}

Table 3.4: Mean absolute difference between the model parameter estimates $(\kappa^*, \ldots, \rho^*)$ from the least squares calibration and the true values $(\kappa^\dagger, \ldots, \rho^\dagger)$ for the FFT method, the COS method, the SC method, and the GMSC method.

3.4.2 Regularization Tests

For the second part of our numerical experiments, we tested the effect of the regularization on the objective function on market data. More specifically, instead of solving (3.4), we solved (3.16) for a data setup similar to that of the non-regularized tests. We performed our numerical tests on S&P500 option data from the 3rd of January 2017 to the 8th of March 2017, for a total of 44 trading days. The total number of options (SPX and SPXW) quoted during this period was 376865.

As in the first part, each calibration consisted of 100 benchmark options; 10 strikes per maturity, for a total of 10 maturities per trading day. The list of maturities consisted of approximately[13] the same maturities as listed for the speed tests. For each maturity, we chose the 5 out-of-the money call options and 5 out-of-the money put options with the highest trading volume. We also included pricing error as an additional stopping criterion. More specifically, we added the condition that change in pricing error should exceed machine precision and that the pricing error should be strictly decreasing with each iteration.

Using this market data, we find signs of apparent non-convexity of the least squares problem to be prevalent when calibrating the 3/2 model. This is in contrast to our tests on simulated data. Another issue is that for many of the trading days, the calibration tends to return implausible parameter estimates when unconstrained. In the case of box constraints, the solu-

13. In the event that less than ten liquid options were present for a given target maturity for a given trading day, we picked the closest maturity instead which fulfilled the liquid option criteria. In no instance was there a trading day in which we did not have exactly ten maturities, with ten strikes per maturity.

tion to (3.4) generally tended to be on one or more of the boundaries. To illustrate this, we solve (3.4) using three different starting points, each of which is close to a different boundary corner. As can be seen in Figure 3.1, the parameter estimates returned by the optimization routine for the different starting points differ greatly, implying that the objective function either contains strict local minima, or is so flat that the optimization fails to find the global optimum due to insufficient precision.

The reason we observe this behavior is due to the interaction between the model and the data. The calibration of option pricing models in the risk neutral world for indices like the S&P500 is dominated by close-to-the-money plain European options because they are the most liquid financial derivatives available. The problem is that the prices of these particular options are less sensitive to changes in the parameters of stochastic volatility models than, say, path-dependent options[14]. This moderate sensitivity leads to an ill-conditioned Hessian of the least squares objective function, which poses numerical problems for a second order gradient-based minimization procedure like the Levenberg-Marquardt method. A consequence of an ill-conditioned Hessian is that noise in the data (see section 3.3 for a discussion on noise in option price data) can lead to outsized effects on the calibration results.

To address this problem, we calibrated the regularized version of the objective function, i.e., (3.16), using the penalty function proposed in section 3.3. As previously mentioned, the regularization parameter χ can be thought to represent how strongly the user of the pricing model believes in the historical estimate, as well as the user's preferences over model risk. In other words, this value cannot be derived objectively a priori. Instead, the value ultimately depends on the real world context in which it is being used.

For illustrative purposes, however, we describe the approach given in Cont and Tankov [2004] which builds on Morozov's discrepancy principle (Morozov [1966]) to calculate χ. It requires that we determine the intrinsic error ϵ_0 in the data we have. In our case that means the lower bound on the quadratic calibration error. An obvious source of intrinsic error in the option price data is the bid-ask spread.

14. While more exotic options would potentially allow for a more numerically stable least squares calibration procedure, such options are generally traded over the counter in the market which makes them too illiquid to yield reliable parameter estimates. Indeed, the main reason we are interested in stochastic volatility models such as the 3/2 model is that they allow us to derive the prices of complex instruments such as path-dependent options in a way that is empirically consistent (by virtue of the model calibration) with the prices of exchange traded options (such as plain European options) that are liquid enough to give an accurate representation of the market pricing measure.

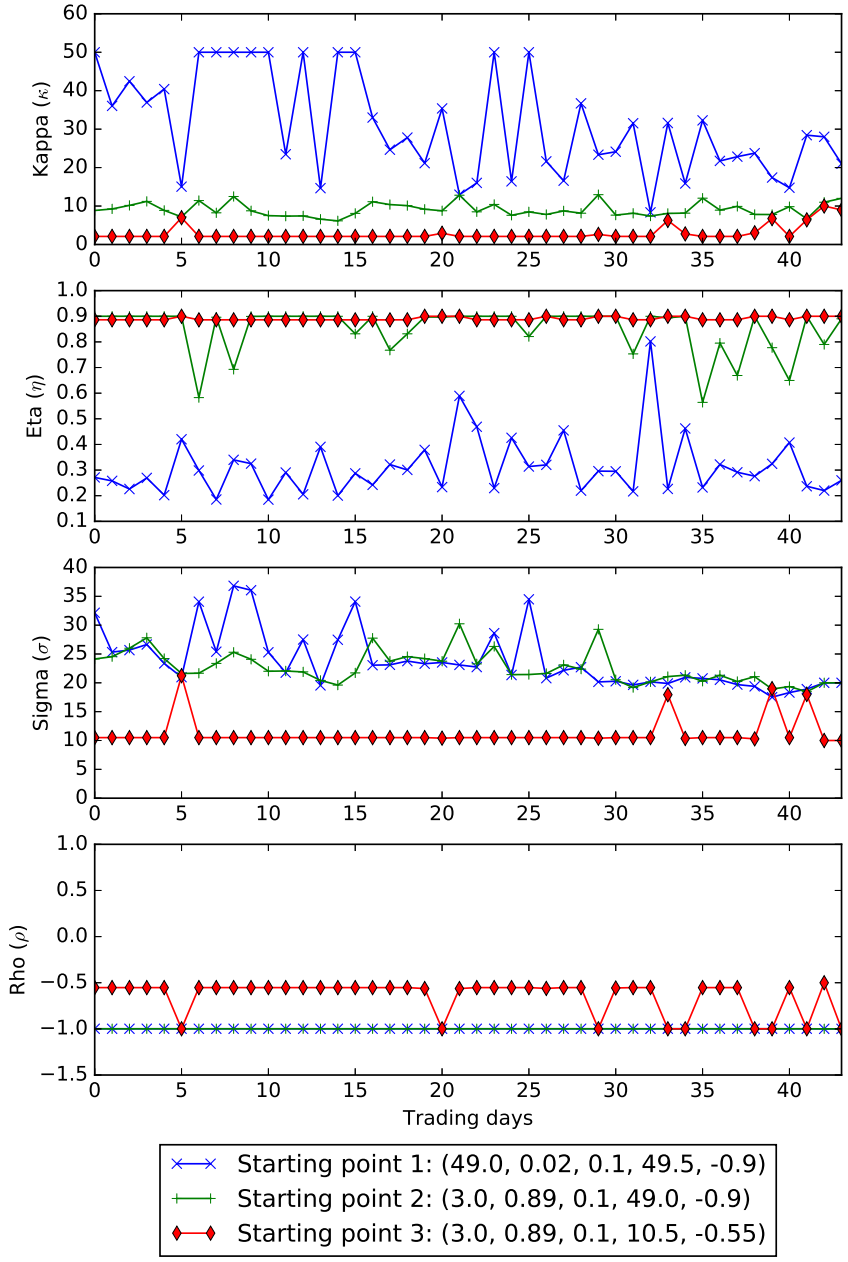

Figure 3.1: Estimated parameter values obtained by solving (3.4) for S&P 500 option data for the first 44 trading days of 2017, using different starting points for the model parameters $\boldsymbol{\Theta} = (\kappa, \eta, \upsilon, \sigma, \rho)$.

Let

$$\epsilon_0^2 = \sum_{i=1}^{N} w_i \left(\pi_i^{\text{Ask}} - \pi_i^{\text{Bid}}\right)^2,$$

where w_i is defined as in (3.4). The basic intuition behind Morozov's discrepancy principle is that we expect that a successful least squares calibration of our model will yield an error

$$\epsilon(\chi)^2 = \min_{\boldsymbol{\Theta}} \left\{ \sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta})^2 + \chi H(\boldsymbol{\Theta} - \tilde{\boldsymbol{\Theta}}) \right\}$$

that is not too far away from ϵ_0^2. The problem then becomes that of solving $\epsilon(\chi) = \delta\epsilon_0$ for χ, where δ is some number larger than 1.

This equation can only be solved through iterative root methods. Each iteration requires solving (3.16), making this a rather expensive operation. For this reason we only solve this equation for the first trading day, and then use the resulting χ^* for the regularized calibration of the remaining days. We find that this generates very stable parameter estimates across trading days, while the calibration error remains close to the error of the non-regularized calibration for most of the trading days. In addition, restarting the calibration from different initial guesses for each trading day reveals that the regularization eliminates any sign of non-convexity.

We performed the initial historical estimation using the history of daily closing prices of the S&P500 index from the 2nd of January 2014 to the 2nd of January 2017 and the VIX from that same period. For the MCMC estimation we used the following parameter priors:

$$\psi \sim N\left(0, \frac{\Omega}{2}\right), \quad \Omega \sim IG\left(2, \frac{1}{200}\right), \quad \kappa \sim N(0,1), \quad \eta \sim N(0,1),$$

where the distributions for κ and η were truncated at zero (making the posteriors also truncated at zero)[15]. These priors derive from those used in Li et al. [2006] and Eraker et al. [2003] for the estimation of the Heston model. The number of samples drawn was 20000, with a burn-in of 5000 iterations, and the total estimation procedure took roughly 18.3 seconds.

For each successive trading day for which we calibrated the 3/2 model we updated the historical estimates by the change in price and VIX value realized on that same day, using 2000 iterations and no burn-in, since the previous estimate acts as an informative prior, as explained in section 3.3. The updating estimation took roughly 1.5 seconds on average.

15. These are uninformative priors and are completely dominated by the likelihood function due to the size of our data sample. As we verified numerically, changing the parameters of the prior distributions appeared to have no discernible effect on the historical estimates.

The resulting historical estimates, $\tilde{\boldsymbol{\Theta}}$ and the corresponding covariance matrix Q for those estimates are given respectively by

$$\tilde{\boldsymbol{\Theta}} \equiv \begin{bmatrix} \kappa \\ \eta \\ \sigma \\ \rho \end{bmatrix} = \begin{bmatrix} 22.46 \\ 0.4947 \\ 22.72 \\ -0.9085 \end{bmatrix},$$

and

$$\boldsymbol{Q} = \begin{bmatrix} 2.702 & -0.007951 & 0.02372 & -0.0002194 \\ -0.007951 & 0.0001007 & -0.0003903 & 0.000004462 \\ 0.02372 & -0.0003903 & 0.3577 & -0.003051 \\ -0.0002194 & 0.000004462 & -0.003051 & 0.00002612 \end{bmatrix}.$$

We used the bisection root finding method for calculating χ^*, with the bisection terminating once a search interval of length 0.0001 was reached. We set the initial interval for χ as $[0,5]$, and the calibration weight, w_i, for option i to the reciprocal of its bid-ask spread, i.e.,

$$w_i = \frac{1}{\pi_i^{\text{Ask}} - \pi_i^{\text{Bid}}},$$

for $i = 1,\ldots,I$. The corresponding calibration results are shown in Figures 3.2 and 3.3.

The results in Figure 3.2 show that the parameter estimates obtained from the regularized calibration are far more stable across trading days than the estimates we obtain from the unregularized calibration. Furthermore, starting the regularized calibration from the different initial points described in the caption of Figure 3.1 all lead to the same solution, meaning that we could not detect any non-convexity with the objective function of the regularized calibration.

The results in Figure 3.3 show that the increased parameter stability is accompanied on average by a very modest decrease in the in-sample fit, as measured by the normalized mean squared absolute price error as well as the root mean squared implied volatility error. Overall, the trade-off between parameter stability and in-sample fit of the regularized version appears very favorable, making clear the usefulness of the regularization approach.

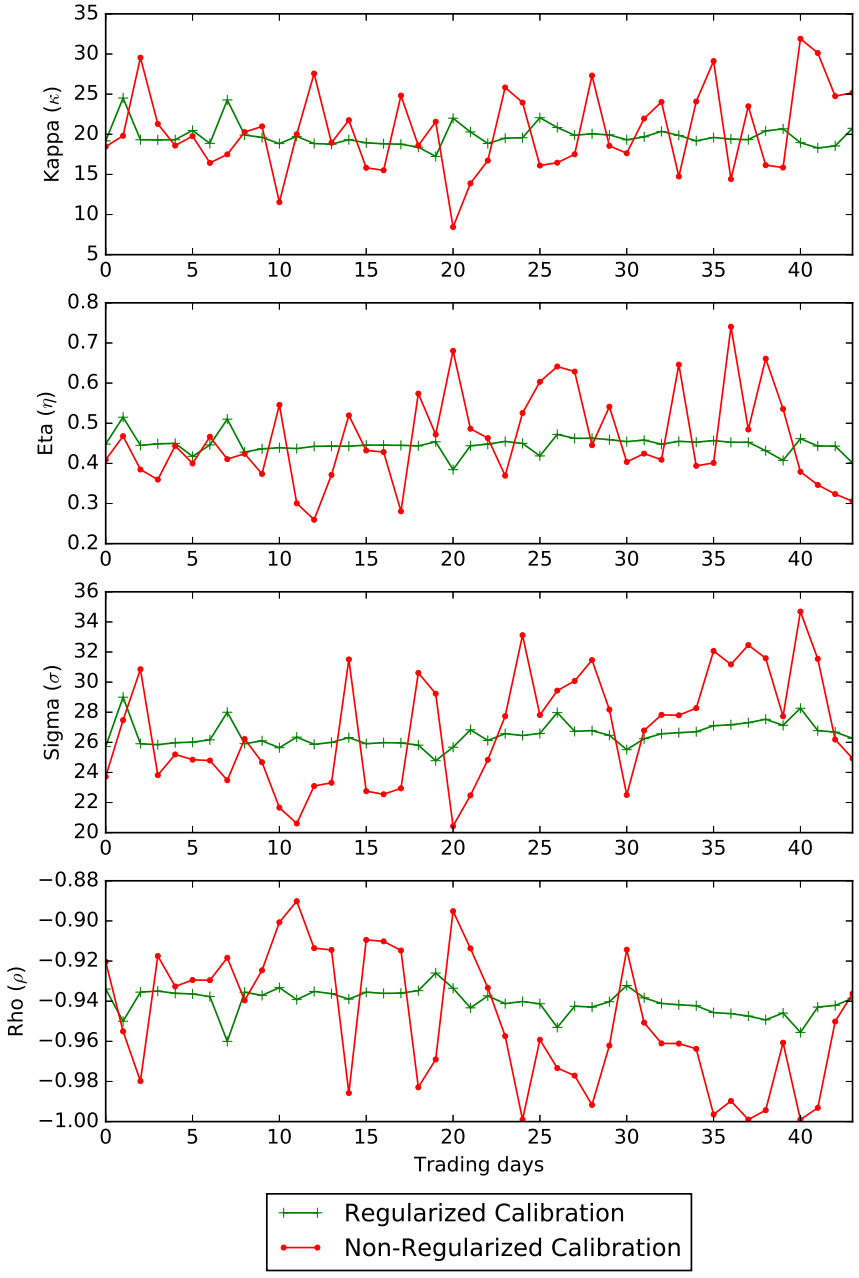

Figure 3.2: Comparison of parameter stability for regularized ($\chi = 3.86$) and non-regularized calibration of the 3/2 model.

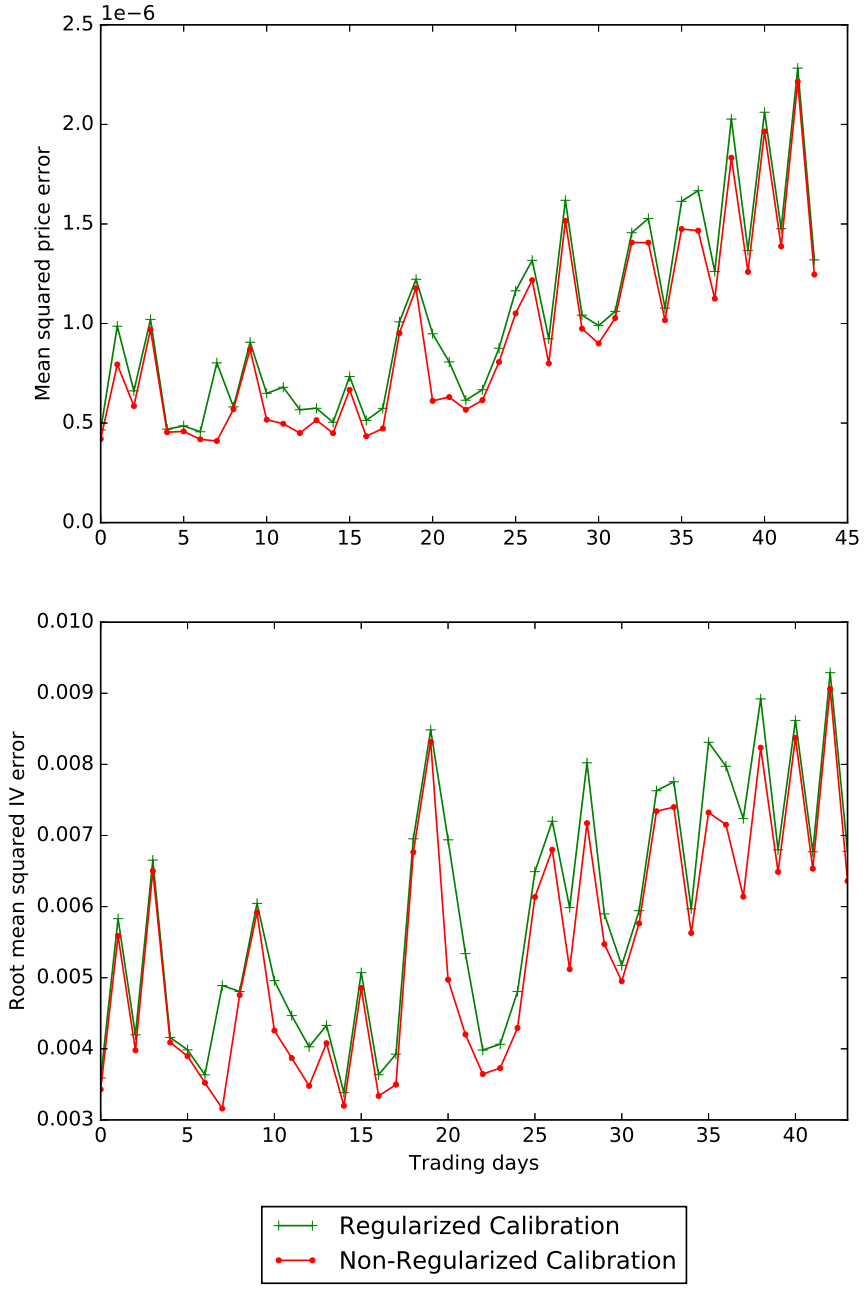

Figure 3.3: Comparison of in-sample calibration error between regularized ($\chi = 3.86$) and non-regularized calibration of the 3/2 model.

3.4.3 Calibration Of Multi-Factor Models

In this section, we show how the GMSC method for the 3/2 model can be used when the model is coupled with other stochastic factors for which we know the characteristic function in closed form. We begin by recalling a well known fact from characteristic function pricing. For the purpose of demonstration, consider the following general multi-factor log-price model,

$$dX_t = \sum_{m=1}^{M} dZ_{t,m}, \quad 0 \le t \le T, \tag{3.20}$$

where $Z_{t,m}$ is an adapted semimartingale characterized by the parameter set $\boldsymbol{\Theta}^{(m)}$ such that for each m, the characteristic function for the density of $Z_{T,m}$ is known in closed form and given by $\phi_m\left(\boldsymbol{\Theta}^{(m)}; u\right)$. If the factors $Z_{t,m}$ and $Z_{t,m'}$ are stochastically independent for any m and m' such that $m \neq m'$, then we have that the characteristic function for X_T is known in closed form and is given by

$$\phi\left(\boldsymbol{\Theta}; u\right) = \prod_{m=1}^{M} \phi_m\left(\boldsymbol{\Theta}^{(m)}; u\right), \tag{3.21}$$

where $\boldsymbol{\Theta} = \left[\boldsymbol{\Theta}^{(1)}, \ldots, \boldsymbol{\Theta}^{(M)}\right]$. Calibrating (3.20) through least squares minimization using gradient-based methods means calculating the equivalent of

$$\nabla_{\boldsymbol{\Theta}}\phi = \left[\frac{\phi}{\phi_1}\nabla_{\boldsymbol{\Theta}^{(1)}}\phi_1, \ldots, \frac{\phi}{\phi_M}\nabla_{\boldsymbol{\Theta}^{(M)}}\phi_M\right]. \tag{3.22}$$

Therefore, the gradient can be calculated separately for each factor while minimizing the least squares objective function, which means that the computational burden of the factors that correspond to a 3/2 volatility specification can be efficiently reduced using the results given in this chapter. Furthermore, this can be done while simultaneously using efficient gradient calculation techniques developed for other factors nested in the model. The method given in Cui et al. [2017], for example, can be used here for any of the M factors that correspond to the square root volatility specification of the Heston model.

We conclude our discussion with a numerical study of the efficiency of the gradient calculation methods in this chapter for a double 3/2 model (D32),

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_{t,1}} S_t dB_{t,1} + \sqrt{V_{t,2}} S_t dB_{t,2}, \\ dV_{t,1} &= \kappa_1(\eta_1 V_{t,1} - V_{t,1}^2)dt + \sigma_1 V_{t,1}^{3/2} dW_{t,1}, \\ dV_{t,2} &= \kappa_2(\eta_2 V_{t,2} - V_{t,2}^2)dt + \sigma_2 V_{t,2}^{3/2} dW_{t,2}, \end{aligned} \tag{3.23}$$

and a mixed Heston-3/2 model (MH32),

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{V_{t,1}} S_t dB_{t,1} + \sqrt{V_{t,2}} S_t dB_{t,2}, \\ dV_{t,1} &= \kappa_1(\eta_1 V_{t,1} - V_{t,1}^2)dt + \sigma_1 V_{t,1}^{3/2} dW_{t,1}, \\ dV_{t,2} &= \kappa_2(\eta_2 - V_{t,2})dt + \sigma_2 \sqrt{V_{t,2}} dW_{t,2}, \end{aligned} \tag{3.24}$$

where W_1, W_2, B_1 and B_2 are standard Brownian motions such that B_1 and W_1 have instantaneous correlation ρ_1, and B_2 and W_2 have instantaneous correlation ρ_2, with all other factor correlations equal to zero.

The setup of our numerical tests on D32 and MH32 is the same as for the simulated gradient and calibration speed tests for the one factor 3/2 model in section 3.4.1, with the exception that we focus solely on the SC method of Kilin and the GMSC. In addition, the true values for $\kappa_2, \eta_2, \nu_2, \sigma_2, \rho_2$ for the square root volatility process in MH32 are chosen randomly from within the box constraints given in Table 3.5.

	min	max
κ_2	0.1	10.0
η_2	0.01	0.90
ν_2	0.01	0.90
σ_2	0.1	1.0
ρ_2	-1.0	0.50

Table 3.5: Variable Constraints. These represent the intervals within which we constrained the generation of random parameter values for the square root volatility component of MH32. All other parameter values were simulated using the box constraints given in Table 3.1.

The numerical results are given in Tables 3.6 and 3.7, for the gradient computations and calibration tests, respectively.

For our tests on D32 we use the GMSC with the analytic gradient derived in this chapter separately for each 3/2 factor. More specifically, we compute (3.22) by applying Algorithm 2 separately to $\nabla_{\boldsymbol{\Theta}^{(1)}}\phi_1$, and $\nabla_{\boldsymbol{\Theta}^{(2)}}\phi_2$, where ϕ_1 and ϕ_2 are the characteristic functions for single factor 3/2 volatility models with $\boldsymbol{\Theta}^{(1)} = \left[\kappa_1, \eta_1, \nu_1, \sigma_1, \rho_1\right]$ and $\boldsymbol{\Theta}^{(2)} = \left[\kappa_2, \eta_2, \nu_2, \sigma_2, \rho_2\right]$, respectively.

For MH32 we denote with ϕ_1 the characteristic function for the one factor 3/2 volatility model, and with ϕ_2 the characteristic function for the one factor square root volatility model. We compute (3.22) by applying Algorithm 2

to $\nabla_{\boldsymbol{\Theta}^{(1)}}\phi_1$, and by applying the gradient computation formulas presented in Cui et al. [2017] for the Heston model to $\nabla_{\boldsymbol{\Theta}^{(2)}}\phi_2$. We refer to this combined method as the mixed gradient-maturity-strike caching (MGMSC) method in Tables 3.6 and 3.7.

	SC D32	GMSC D32	SC MH32	MGMSC MH32
mean	1.231	0.08375	0.6502	0.04807
std	0.8012	0.03913	0.5824	0.02796
max	6.429	0.1799	4.313	0.1271
min	0.9140	0.0415	0.4040	0.01668

Table 3.6: Computational times in seconds for the analytic gradient computations with the GMSC method for D32, and with the GMSC coupled with the formula from Cui et al. [2017] (MGMSC) for MH32. The computational times in seconds for the central finite difference gradient approximation using SC are given as comparison for both models.

	SC D32	GMSC D32	SC MH32	MGMSC MH32
mean	701.2	58.18	95.67	8.302
std	665.0	47.85	79.15	6.224
max	2374	186.6	325.0	29.08
min	193.6	9.077	10.12	1.257

Table 3.7: Computational times in seconds for the calibration runs using the analytic gradient with GMSC method for D32, and the GMSC coupled with the formula from Cui et al. [2017] (MGMSC) for MH32. The computational times in seconds for the calibration runs using the central finite difference gradient approximation and SC are given as comparison for both models.

As expected, the time it takes to compute the gradient for D32 is roughly twice the time it takes for calculating the gradient for the single factor 3/2 model for both the SC and the GMSC methods. In contrast, the computational cost of the calibration procedure for the multifactor models is significantly higher. This is to be expected as well, since the iteration count of the Newton-based minimization procedure grows superlinearly with the number of decision variables. This serves to highlight the necessity for numerically efficient calibration techniques for multi-factor models.

3.5 Conclusion

In this chapter we have derived the analytic gradient for the characteristic function of the 3/2 model of stochastic volatility and devised an algorithm that exploits its mathematical features to avoid redundant calculations when calibrating the 3/2 model to option data using the standard nonlinear least squares approach. We have shown how this can greatly speed up the calibration process. In addition, we have proposed a form of regularization for the calibration problem that uses MCMC estimation on historical data to produce both a regularization point, as well as a damping matrix which we use to produce a parsimonious L_2 penalty function.

In our numerical experiments, we compared our calibration algorithm to the FFT method of Carr and Madan, the COS method of Fang and Oosterlee, and the SC method of Kilin. We find that the method presented here outperforms these by a factor of roughly 37, 12 and 11, respectively. For researchers or practitioners looking to calibrate the 3/2 model, the method presented here is, to our knowledge, the fastest way to do it so far presented in the literature.

Our numerical experiments furthermore demonstrated that for the market data we had at our disposal, the least squares calibration of the 3/2 model leads to highly unstable parameter estimates, both with respect to the initial point of the optimization procedure, as well as across trading days. The regularization method proposed in this chapter was shown to be an effective tool to deal with these issues, producing a very favorable ratio of parameter estimation stability to in-sample fit.

One aspect of characteristic function pricing which we have not delved into here is the issue of discontinuities due to branch cuts in the complex plane. This problem has attracted considerable attention in the case of the Heston model (see e.g. Cui et al. [2017]; del Baño et al. [2010]; Albrecher et al. [2007]; Jäckel and Kahl [2005]). However, it remains an open problem in the case of the 3/2 model, and is of interest with respect to further study and refinement of this model.

Chapter 4

Non-Affine Stochastic Volatility with Seasonal Trends

4.1 Introduction

Certain classes of commodity derivatives, such as futures written on agricultural products and natural gas, are known to exhibit predictable cyclical trends with respect to calendar time (Fama and French [1987], Choi and Longstaff [1985], Sørensen [2002], Back et al. [2013]). This has motivated researchers to consider modified option pricing models for pricing options on futures contracts on those particular commodities that take into account seasonal patterns. Seasonality in the price of the underlying should have no effect on the option pricing dynamics under the risk neutral measure by an argument of no-arbitrage, while seasonality in volatility needs to be addressed. Indeed, in Figures 4.1 and 4.2 we see that the implied volatility indexes for options on futures contracts for corn and wheat tend to demonstrate seasonal highs and lows[16].

In Koekebakker and Lien [2004] a jump-diffusion model is proposed, where the volatility function for the diffusion term is deterministic but cyclical in calendar time. In Richter and Sørensen [2002] a stochastic volatility model is proposed that includes a convenience yield process and deterministic seasonal effects. A model based on seasonally modified Ornstein-Uhlenbeck processes is proposed in Back et al. [2013]. In Arismendi et al. [2016] a modified Heston model is proposed where the long run mean captures seasonal

16. Figure 4.1 shows daily closing spot values from the CBOE/CBOT Corn Volatility Index for the period 1st of October 2011 to the 1st of June 2018, and Figure 4.2 shows the daily closing spot values from the CBOE/CBOT Wheat Volatility Index for the period of 1st of September 2012 to the 1st of June 2018.

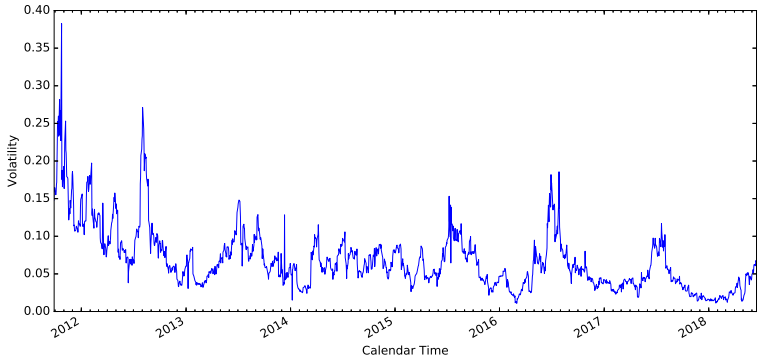

Figure 4.1: Corn Implied Volatility Index.

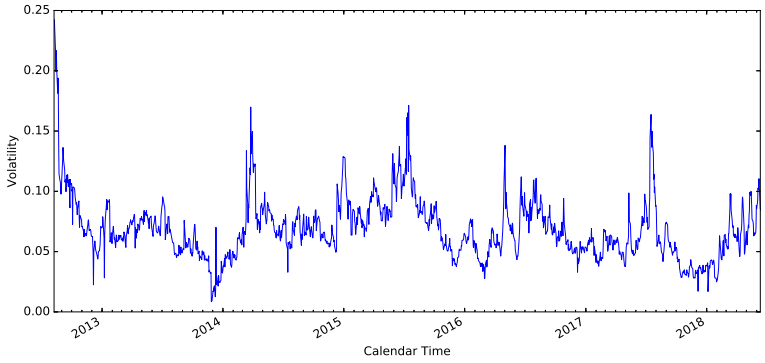

Figure 4.2: Wheat Implied Volatility Index.

variation of the volatility of the underlying, along with a MCMC-based estimation procedure for a robust calibration of the model. In Schneider and Tavin [2015] a modification of the Heston model is proposed which accommodates both the Samuelson effect, as well as seasonal trends in volatility.

These models all share one commonality; they include purely affine volatility dynamics. In this chapter, however, we propose a new, non-affine stochastic volatility model that takes seasonal effects into account. More specifically, we propose a generalization of the 3/2 stochastic volatility model which includes an explicit seasonal volatility component, for which we derive the characteristic function of the model density in closed form.

The rest of the chapter is as follows. In section 4.2 we present the new model and its characteristic function. In section 4.3 we go through the model calibration methodology used in our numerical tests. And in section 4.4 we present and discuss the results from our numerical tests on the proposed model using market prices of options on corn and wheat futures.

4.2 Non-Affine Dynamics With Seasonally Varying Volatility Trend

Today, participation in commodity markets generally does not entail buying and selling the commodities directly, but rather the futures contracts written on those commodities. Consequently, option contracts in these markets are generally written on the futures. We therefore treat the futures contracts as the underlying assets in our study of option pricing for commodities, with the aim of capturing their empirical features in our option pricing model. With that in mind, the 3/2 model in its most basic form under the physical (i.e., real world) measure for the price process of a futures contract is characterized by

$$\begin{aligned} dF_t(T) &= \sqrt{V_t} F_t(T) dB_t, \\ dV_t &= \kappa(\eta V_t - V_t^2)\,dt + \sigma {V_t}^{3/2} dW_t, \end{aligned} \tag{4.1}$$

where $F_t(T)$ is the price of a futures contract with expiration date T at time t, and B and W are standard Brownian motions with an instantaneous correlation coefficient given by ρ, i.e., $\mathrm{Cov}(dB_t, dW_t) = \rho\,dt$.

We can see from the dynamics of the 3/2 model, however, that the model is homogeneous with respect to calendar time. This is a clear disadvantage for modeling assets like the futures market on corn which show clear seasonal trends. This motivates us to consider an alternative specification, where the dynamics include an explicit function of calendar time.

With that in mind, we turn to the main contribution of this chapter, which is a 3/2 stochastic volatility model where the volatility trend incorporates seasonal fluctuations. For this model we derive the characteristic function in closed form. The dynamics of the model we propose are given by

$$\begin{aligned} dF_t(T) &= \sqrt{V_t} F_t(T) dB_t, \\ dV_t &= \kappa\left(\chi(t) V_t - V_t^2\right) dt + \sigma {V_t}^{3/2} dW_t, \\ \chi(t) &= \eta + \lambda \left| \frac{1}{2} - (t - \tau - \lfloor t - \tau \rfloor) \right|, \end{aligned} \tag{4.2}$$

where B and W are as in (4.1), τ is the calendar time at which the long run volatility is at its highest, λ corresponds to the strength of the seasonal effect, and $\lfloor \cdot \rfloor$ is the integer floor function, i.e., $\lfloor t \rfloor = \max\{m \in \mathbb{Z} \mid m \leq t\}$. The corresponding PDE implied by the Feynman-Kac theorem is given by

$$\frac{\partial J}{\partial t} + \frac{VF^2}{2}\frac{\partial^2 J}{\partial F^2} + \sigma\rho F V^2 \frac{\partial^2 J}{\partial V \partial F} + \frac{\sigma^2 V^3}{2}\frac{\partial^2 J}{\partial V^2} + \kappa V(\chi(t) - V)\frac{\partial J}{\partial V} - rJ = 0. \tag{4.3}$$

Here, the constant long-run volatility parameter in (4.1) is replaced by a periodic triangular-shaped function $\chi(t)$. This function captures the seasonally deterministic trend that we observe in the data for commodities like corn and wheat, where the implied volatility of options on futures with maturities that coincide with the harvest season tends to be higher than that of options that expire immediately before it. Figures 4.3 and 4.4 show the triangular pattern for the mean seasonal volatility parameter estimates for options on corn and wheat futures, respectively, which are presented in section 4.4.4.

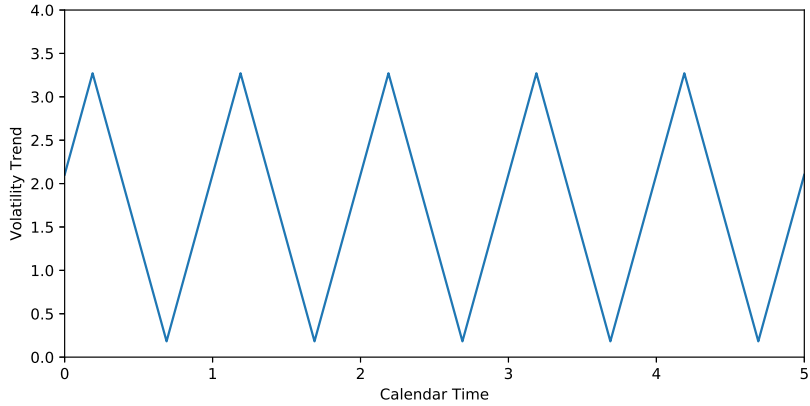

Figure 4.3: Seasonal volatility component $\chi(t)$ with the mean parameter estimates for options on corn futures as presented in Table 4.2 ($\eta = 0.182$, $\lambda = 6.179$ and $\tau = 0.189$).

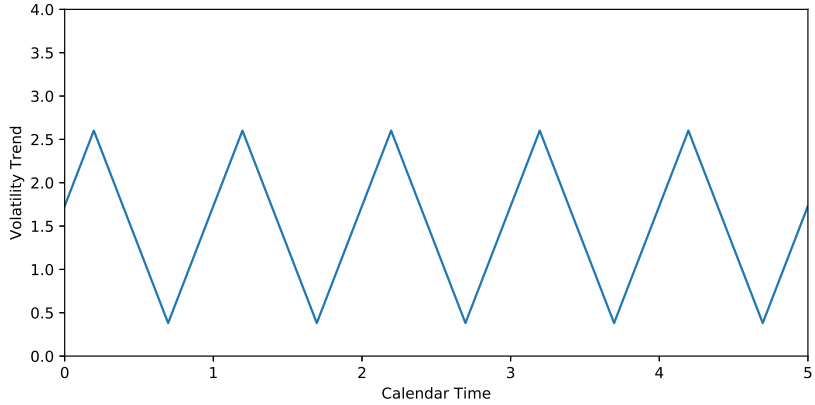

Figure 4.4: Seasonal volatility component $\chi(t)$ with the mean parameter estimates for options on wheat futures as presented in Table 4.3 ($\eta = 0.381$, $\lambda = 4.441$ and $\tau = 0.196$).

As is the case with the standard 3/2 model, the probability density of F_T in (4.2) is not known explicitly in closed form. On the other hand, we can derive the characteristic function[17] $\phi(\mathbf{\Theta}, T; u) \equiv \mathbb{E}^{\mathbb{Q}}\left[e^{iuX_T}\right]$ of $X_T \equiv \ln(F_T)$. The formula for $\phi(\mathbf{\Theta}, T; u)$ is given in Theorem 4.1.

Theorem 4.1. *Let F_t for $0 \leq t \leq T$ be defined as in (4.2), and let $x \equiv \ln(F_0)$. Then the characteristic function of X_T is given by*

$$\phi(\mathbf{\Theta}, T; u) = e^{iux}\frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\zeta^{\alpha}M\left(\alpha;\beta;-\zeta\right), \tag{4.4}$$

where $M(\alpha;\beta,-\zeta)$ is the confluent hypergeometric function of the first kind,

$$M(\alpha;\beta;-\zeta) \equiv \sum_{n=0}^{\infty}\frac{(\alpha)_n}{(\beta)_n}\frac{(-\zeta)^n}{n!}, \tag{4.5}$$

with

$$(x)_n \equiv \begin{cases}\prod_{i=0}^{n-1}(x+i) & \text{if } n>0,\\ 1 & \text{if } n=0\end{cases}$$

and α, β, and ζ are given by

$$\begin{aligned}
\zeta &\equiv \frac{2}{v\sigma^2\Omega},\\
\alpha &\equiv -\left(\frac{1}{2}-\frac{p}{\sigma^2}\right)+\sqrt{\left(\frac{1}{2}-\frac{p}{\sigma^2}\right)^2+\frac{2q}{\sigma^2}},\\
\beta &\equiv 2\left[\alpha+1-\frac{p}{\sigma^2}\right],\\
p &\equiv -\kappa+i\sigma\rho u,\\
q &\equiv \frac{iu}{2}+\frac{u^2}{2},
\end{aligned} \tag{4.6}$$

17. Without loss of generality, we take the initial time to be 0 throughout the chapter to simplify notation, and take the expiration date of the underlying futures contract to be given as T, writing F_t instead of $F_t(T)$.

where

$$
\begin{aligned}
\Omega \equiv \sum_{m=-1/2}^{M} e^{\xi_m} \sqrt{-\frac{\pi}{2\gamma_m\kappa\lambda}} \Bigg(& erf\left(\sqrt{\frac{-\gamma_m\kappa\lambda}{2}}\left(m-\lfloor m\rfloor+\gamma_m\frac{\eta}{\lambda}\right)\right) \\
& - erf\left(\sqrt{\frac{-\gamma_m\kappa\lambda}{2}}\left(m-\lfloor m\rfloor-\frac{1}{2}+\gamma_m\frac{\eta}{\lambda}\right)\right)\Bigg) \\
+ e^{\xi_M}\sqrt{\frac{-\pi}{2\gamma_M\kappa\lambda}}\Bigg(& erf\left(\sqrt{\frac{-\gamma_M\kappa\lambda}{2}}\left(T-\tau-\lfloor M\rfloor-\frac{1}{2}+\gamma_M\frac{\eta}{\lambda}\right)\right) \\
& - erf\left(\sqrt{\frac{-\gamma_M\kappa\lambda}{2}}\left(M-\lfloor M\rfloor-\frac{1}{2}+\gamma_M\frac{\eta}{\lambda}\right)\right)\Bigg) \\
+ e^{\xi_{-\frac{1}{2}}}\sqrt{\frac{-\pi}{2\kappa\lambda}}\Bigg(& erf\left(\sqrt{\frac{-\kappa\lambda}{2}}\left(\frac{\eta}{\lambda}-2\right)\right) - erf\left(\sqrt{\frac{-\kappa\lambda}{2}}\left(-\tau-\frac{3}{2}+\frac{\eta}{\lambda}\right)\right)\Bigg), \\
\xi_m \equiv \left(\tau-\frac{1}{2}\right)\left(\kappa\eta+\frac{1}{2}\lambda\left|\tau-\frac{1}{2}\right|\right) & + \left(m+\frac{1}{2}\right)\left(\eta+\frac{\kappa\lambda}{4}\right) - \kappa\eta(\tau+m) \\
& - \frac{\gamma_m\kappa\lambda}{2}(\tau+m)^2 - \left(\tau+\lfloor m\rfloor+\frac{1}{2}\right)(\tau+m) \\
& - \frac{\gamma_m\kappa\lambda}{2}\left(\tau+\lfloor m\rfloor+\frac{1}{2}-\gamma_m\frac{\eta}{\lambda}\right)^2, \\
erf(x) \equiv \int_0^x e^{-t^2}dt, & \\
M \equiv \frac{\lfloor 2T-2\tau\rfloor}{2}, & \\
\gamma_m \equiv (-1)^{2m+1}. &
\end{aligned}
\tag{4.7}
$$

Proof. From Carr and Sun [2007] we know that the characteristic function for the stochastic process defined by

$$
\begin{aligned}
dX_t &= \mu X_t dt + \sqrt{V_t}X_t dB_t, \\
dV_t &= \kappa\left(\chi(t)V_t - V_t^2\right)dt + \sigma {V_t}^{3/2} dW_t,
\end{aligned}
\tag{4.8}
$$

where $\chi(t)$ is an arbitrary deterministic function of time, then the characteristic function of $\ln(X_T)$ is given by

$$
\phi(\mathbf{\Theta}, T; u) = e^{iux}\frac{\Gamma(\beta-\alpha)}{\Gamma(\beta)}\zeta^{\alpha}M\left(\alpha;\beta;-\zeta\right), \tag{4.9}
$$

where

$$\begin{aligned}\zeta &\equiv \nu\frac{2}{\nu\sigma^2\Omega(t)},\\ \Omega(t) &\equiv \int_0^T e^{\int_0^u \kappa\chi(s)ds}du,\end{aligned} \tag{4.10}$$

and where $M(\alpha;\beta,-\zeta)$, α,and β are defined as in (4.6). Setting

$$\chi(t) = \eta + \lambda\left|\frac{1}{2} - (t-\tau-\lfloor t-\tau\rfloor)\right|$$

we have that

$$\begin{aligned}\int_0^u \kappa\chi(s)ds &= \int_0^\tau \kappa\chi(s)ds + \int_\tau^m \kappa\chi(s)ds + \int_m^u \kappa\chi(s)ds\\ &= \left(\tau-\frac{1}{2}\right)\left(\kappa\eta + \frac{\kappa\lambda}{2}\left|\tau-\frac{1}{2}\right|\right) + \left(m+\frac{1}{2}\right)\left(\kappa\eta+\frac{\kappa\lambda}{4}\right)\\ &\quad + \kappa\eta u + (-1)^{2m+1}\kappa\lambda\left(\frac{u^2}{2} - \left(\tau+n+\frac{1}{2}\right)u\right)\\ &\quad - \left(\kappa\eta(\tau+m) + (-1)^{2m+1}\kappa\lambda\left(\frac{(\tau+m)^2}{2} - \left(\tau+n+\frac{1}{2}\right)(\tau+m)\right)\right),\end{aligned}$$

where $m = \frac{1}{2}\lfloor 2u-2\tau\rfloor$, and $n = \lfloor u-\tau\rfloor$. Writing

$$g(u) = \int_0^u \kappa\chi(s)ds,$$

we want to find

$$A_1 = \int_0^\tau e^{g(u)}du, A_2 = \int_\tau^M e^{g(u)}du, \text{ and } A_3 = \int_M^T e^{g(u)}du.$$

Let $\gamma(m) = (-1)^{2m+1}$ and

$$\begin{aligned}\xi(m) &= \left(\tau-\frac{1}{2}\right)\left(\kappa\eta + \frac{1}{2}\kappa\lambda\left|\tau-\frac{1}{2}\right|\right) + \left(m+\frac{1}{2}\right)\left(\kappa\eta+\frac{\kappa\lambda}{4}\right) - \kappa\eta(\tau+m)\\ &\quad - \gamma(m)\kappa\lambda\left(\frac{1}{2}(\tau+m)^2 - \left(\tau+n+\frac{1}{2}\right)(\tau+m)\right)\\ &\quad - \frac{1}{2}\gamma(m)\kappa\lambda\left(\tau+n+\frac{1}{2} - \gamma(m)\frac{\eta}{\lambda}\right)^2.\end{aligned}$$

Then we have that for $0 \le u \le \tau$ we can write

$$g(u) = \xi\left(-\frac{1}{2}\right) + \frac{1}{2}\gamma(m)\kappa\lambda\left(u-\tau-\frac{3}{2}+\gamma(m)\frac{\eta}{\lambda}\right)^2, \tag{4.11}$$

which means that we can write

$$A_1 = e^{\xi\left(-\frac{1}{2}\right)}\sqrt{-\frac{\pi}{2\kappa\lambda}}\left(\operatorname{erf}\left(\sqrt{-\frac{\kappa\lambda}{2}}\left(\frac{\eta}{2\lambda}-3\right)\right)-\operatorname{erf}\left(\sqrt{-\frac{\kappa\lambda}{2}}\left(\frac{\eta}{2\lambda}-\tau-3\right)\right)\right).$$

Similarly, we see that

$$\begin{aligned}A_2 &= \sum_{m=-\frac{1}{2}}^{M}\int_m^{m+\frac{1}{2}} e^{\xi(m)}e^{\frac{1}{2}\gamma(m)\kappa\lambda\left(u-\tau-\lfloor m\rfloor-\frac{1}{2}+\gamma(m)\frac{\eta}{\lambda}\right)^2}du \\ &= \sum_{m=-\frac{1}{2}}^{M}\left\{e^{\xi(m)}\sqrt{-\frac{\pi}{2\kappa\lambda}}\left[\operatorname{erf}\left(\sqrt{-\frac{\gamma(m)\kappa\lambda}{2}}\left(m-\lfloor m\rfloor+\gamma(m)\frac{\eta}{\lambda}\right)\right)\right.\right. \\ &\qquad \left.\left.-\operatorname{erf}\left(\sqrt{-\frac{\gamma(m)\kappa\lambda}{2}}\left(m-\lfloor m\rfloor-\frac{1}{2}+\gamma(m)\frac{\eta}{\lambda}\right)\right)\right]\right\}.\end{aligned}$$

And, finally, we have that

$$\begin{aligned}A_3 =& e^{\xi(M)}\sqrt{-\frac{\pi}{2\gamma(M)\kappa\lambda}}\left\{\operatorname{erf}\left(\sqrt{-\frac{\gamma(M)\kappa\lambda}{2}}\left(T-\tau-\lfloor M\rfloor-\frac{1}{2}+\gamma(M)\frac{\eta}{2\lambda}\right)\right)\right. \\ &\left.-\operatorname{erf}\left(\sqrt{-\frac{\gamma(M)\kappa\lambda}{2}}\left(M-\lfloor M\rfloor-\frac{1}{2}+\gamma(M)\frac{\eta}{2\lambda}\right)\right)\right\}.\end{aligned}$$

□

Although the well-known Gaussian error function, erf(x), is given as an integral it is usually computed using analytic approximations that yield an arbitrary level of accuracy, see e.g. Abramowitz and Stegun [1964].

4.3 A Fast and Robust Calibration Algorithm

We now turn our attention to the least squares calibration of the seasonal 3/2 model. We assume throughout this section that the calibration instruments consist of plain European put and call options[18]. Denoting with F_T the price of the underlying future at time T as in the preceding section we will from now on refer to T as the maturity of the option derived from F, with the expiry of the underlying future itself being immaterial apart from the fact that it is assumed to exceed T. With that in mind we recall that the

18. In section 4.4 we relax this assumption.

price π^{call} of a European call option and the price π^{put} of a European put option at time 0 with strike price K and maturity T are respectively given by

$$\pi^{\text{call}} = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[|F_T - K|^+\right],$$
$$\pi^{\text{put}} = e^{-rT}\mathbb{E}^{\mathbb{Q}}\left[|K - F_T|^+\right],$$

where the superscript $\mathbb{Q}$ refers to the fact that the expectation is taken under the risk-neutral measure, rather than the objective measure.

Given a set of options, $i = 1,\ldots,I$, with observed market prices we characterize option i as the pair $(F,K,T)_i$, where F is the underlying future, K is the strike price, and T is the option maturity. We observe the market price π_i^{Market} for each of these options at time $t = 0$, and wish to infer the parameters of our pricing model using these observed market prices. The least squares calibration objective is given by

$$\min_{\boldsymbol{\Theta}} L(\boldsymbol{\Theta}), \tag{4.12}$$

where

$$L(\boldsymbol{\Theta}) = \sum_{i=1}^{I} w_i f_i(\boldsymbol{\Theta})^2, \tag{4.13}$$

and

$$f_i(\boldsymbol{\Theta}) \equiv \pi(\boldsymbol{\Theta};(F,K,T)_i) - \pi_i^{\text{Market}}.$$

Here, w_i is a weight which determines the influence of option price i on the calibration, and $\pi(\boldsymbol{\Theta};(K,T)_i)$ is the price of option i given by the model using the parameter values $\boldsymbol{\Theta}$. In the case of the homogeneous 3/2 model defined by (4.1) we take the parameter set to be given by $\boldsymbol{\Theta} = \{\kappa,\eta,\upsilon,\sigma,\rho\}$. Note that while the spot volatility $\upsilon = V_t$ is strictly speaking not a parameter of the model, it is unobserved and therefore included as a decision variable when solving (4.12). In the case of the seasonal 3/2 model defined by (4.2), we take the parameter set to be given by $\boldsymbol{\Theta} = \{\kappa,\eta,\upsilon,\sigma,\rho,\lambda,\tau\}$.

As explained in chapter 3, computing the solution to (4.12) requires repeated Fourier inversions of the characteristic function of the model. For the calibration procedure of the seasonal 3/2 stochastic model we use the simple strike-caching method described in section 3.2, where the option price is given by (3.8) and the inversion is done through the caching algorithm of Kilin [2011] which is detailed in subsection 3.2.2.

In section 3.4.2 we showed that the homogeneous 3/2 stochastic volatility model yields a least squares objective function that exhibits sign of non-convexity. A more general version of the model, such as the one introduced here, can therefore be expected to suffer from the same problem.

To address this issue, we augment the gradient-based optimization routine with a metaheuristic search algorithm referred to as Differential Evolution (DE). It initializes by randomly generating a set $\{\boldsymbol{\Theta}^{(1)}, \ldots, \boldsymbol{\Theta}^{(N_c)}\}$ of candidate solutions to the optimization problem. Then, for every candidate solution $\boldsymbol{\Theta}^{(c)}$, it randomly selects three other distinct candidates $\boldsymbol{\Theta}^{(j)}, \boldsymbol{\Theta}^{(k)}$ and $\boldsymbol{\Theta}^{(m)}$ and constructs from these a fourth candidate $\boldsymbol{\Theta}^{(n)} = \boldsymbol{\Theta}^{(j)} + F \times (\boldsymbol{\Theta}^{(k)} - \boldsymbol{\Theta}^{(m)})$, where F is a weight parameter.

Next, an element-wise stochastic crossover takes place between $\boldsymbol{\Theta}^{(c)}$ and $\boldsymbol{\Theta}^{(n)}$. What this means is that for every parameter value $\Theta_j^{(c)}$ in $\boldsymbol{\Theta}^{(c)}$ a value a_j^c is randomly generated by sampling the uniform distribution $U[0,1]$ and if this value exceeds a predetermined threshold value CR (referred to as the crossover probability) then $\Theta_j^{(c)}$ is replaced by the corresponding parameter value $\Theta_j^{(n)}$ in $\boldsymbol{\Theta}^{(n)}$. The new candidate solution $\hat{\boldsymbol{\Theta}}^{(c)}$ constructed by performing this crossover for each of the parameter values in $\boldsymbol{\Theta}^{(c)}$ is then compared to the original solution $\boldsymbol{\Theta}^{(c)}$, and if it results in a smaller error it replaces the original solution. Otherwise, $\boldsymbol{\Theta}^{(c)}$ is retained. The current iteration ends once this procedure has been carried out for all the candidate solutions present at the start of the iteration, and the next iteration begins, repeating the mixture-crossover process for each candidate solution. The algorithm terminates either when a solution has been found that yields a sufficiently low calibration error, or it has exhausted its allocated computational budget. Pseudo-code for this procedure is given in Algorithm 4, while a more comprehensive description of DE can be found in Price and Storn [1997].

In our implementation, the best solution generated by the DE algorithm at its termination is subsequently improved by a Quasi-Newton method. As is typical for global optimization algorithms, the introduction of the DE algorithm in the context of solving (4.12) requires a considerably higher number of price computations which includes the rather costly Fourier inversion of the characteristic function of the seasonal 3/2 model. To prevent the calibration procedure becoming too slow, we take advantage of the fact that the Fourier inversion can be computed numerically in parallel, which extends naturally to the individual price computations embedded in the least squares calibration objective function.

The approach we take here mixes the efficient characteristic function computations using the strike caching method described above with the parallelization of the computation of both (4.13) and (3.14) to quickly evaluate the least squares calibration objective function. The computing platform we chose for the parallel computation task is CUDA. Two hardware-related issues guide the implementation[19]. One of these is the performance

19. An in-depth discussion of the technical details of GPU programming and the CUDA API is

bottleneck of reading from and writing to global memory. Efficiently segmenting the execution of the algorithm between the host system and the GPU is therefore important for the overall performance. Another bottleneck is encountered when the results from different threads need to be repeatedly combined since this requires extensive thread synchronization (i.e., for threads to wait on each other) which slows down the overall execution. Consequently, purely parallelizable code should be executed without interruption on the GPU device.

In summary, we implemented the following method. To start with, the specifications of the benchmark options (i.e., the strike price, maturity, and observed market price for each benchmark option) along with the price of the underlying asset and an initial set of model parameter values $\{\boldsymbol{\Theta}^{(c)}\}_{c=1}^{N_c}$ are loaded into global memory[20]. Next, the CUDA kernel is launched with N_c threadblocks, each with N_b threads. The implementation assumes that $N_b = J \times S$, where J is the (fixed) number of integration nodes per option maturity, and S is the number of maturities present among the benchmark options. Thread p in threadblock c first computes $\hat{\phi}_p = \phi\left(\boldsymbol{\Theta}^{(c)}, T_{p'}; u_p\right)$ where $p' = \lfloor \frac{p}{J} \rfloor + 1$ before computing

$$\left\{ \frac{\left(\mathrm{Re}(\hat{\phi}_p) + \frac{\mathrm{Im}(\hat{\phi}_p)}{u_p}\right)}{1+u_p^2}, \frac{\left(\mathrm{Im}(\hat{\phi}_p) - \frac{\mathrm{Re}(\hat{\phi}_p)}{u_p}\right)}{1+u_p^2} \right\}, \tag{4.14}$$

which it stores in the shared memory of p's threadblock. This is followed by thread synchronization, after which thread $p = 1, \ldots, I$ is assigned the task of fetching (4.14) from shared memory and the specifications of option contract number p from global memory to compute (3.8), and writing the weighted least squares error $w_p r_p(\boldsymbol{\Theta})^2$ again to an array in shared memory[21]. Finally, the first thread in the block fetches and sums together the elements $w_1 r_1(\boldsymbol{\Theta})^2, \ldots, w_I f_i(\boldsymbol{\Theta})^2$ computed previously by the selected I threads, and writes the result to global memory.

Assuming the DE routine is executed with N_c candidate solutions, a single call to the CUDA pricing kernel described above launches N_c, each with N_b threads. Once the DE routine has terminated, the gradient-based local search takes over using the output of the DE routine as the initial guess.

beyond the scope of this text, but a comprehensive introduction to these can be found at `https://docs.nvidia.com/cuda/cuda-c-programming-guide/`

20. More specifically, these are loaded into the constant segment of global memory in our implementation.

21. Note that while we start the indexes in Algorithm 3 at 1 to maintain consistency in our presentation, both the block index and the thread index in the CUDA API start at 0, which is the arrangement we use in the actual software implementation of the kernel.

Assuming that a central finite difference approximation is used (as we did in our numerical experiments) to approximate the gradient, a model with N_p parameters will require $2 * N_p$ evaluations of the objective function for a single evaluation of the gradient. In addition, an extra evaluation of the objective function is required for keeping track of the progress of the routine. So, a single iteration of the local search calls the CUDA pricing kernel once, launching a total of $2N_p + 1$ threadblocks, each with N_b threads.

Algorithm 3 CUDA Pricing Kernel

procedure PARALLELPRICINGKERNEL($\{\boldsymbol{\Theta}_c\}_{c=1}^{N_c}, \{(F,K,T)_i\}_{i=1}^{I}$)
  %Shared memory arrays are prefixed with s_m%
  %NodeIndexer(p) calculates (or fetches) element p from $\boldsymbol{U}$%
  %NodeWeighting(p) calculates (or fetches) element p from $\boldsymbol{B}$%
  p $\leftarrow$ threadID
  c $\leftarrow$ blockID
  $u_p \leftarrow$ NodeIndexer(p)
  s_m1[p] $\leftarrow u_p$
  s_m2[p] $\leftarrow$ NodeWeighting(p)
  $p' \leftarrow \lfloor \frac{p}{J} \rfloor + 1$
  $\hat{\phi}_p \leftarrow \phi\left(\boldsymbol{\Theta}^{(c)}, T_{p'}; u_p\right)$
  s_m3[p] $\leftarrow \frac{\left(\mathrm{Re}(\hat{\phi}_p) + \frac{\mathrm{Im}(\hat{\phi}_p)}{u_p}\right)}{1+u_p^2}$
  s_m4[p] $\leftarrow \frac{\left(\mathrm{Im}(\hat{\phi}_p) - \frac{\mathrm{Re}(\hat{\phi}_p)}{u_p}\right)}{1+u_p^2}$
  -Thread Synchronization-
  if p < I **then**
    Compute (3.14) using stored values in s_m1, s_m2, s_m3 and s_m4
    Compute $\pi\left(\boldsymbol{\Theta}; (K,T)_p\right)$ using (3.8)
    s_m5[p] $\leftarrow w_p r_p(\boldsymbol{\Theta})^2$
  end if
  -Thread Synchronization-
  if p = 1 **then**
    g_m[c] $\leftarrow$ Sum(s_m5)
  end if
end procedure

Algorithm 4 Parallelized Differential Evolution

procedure DE($\{(F,K,T)_i\}_{i=1}^{I}$)
  Randomly Generate $\left\{\boldsymbol{\Theta}^{(1)},\ldots,\boldsymbol{\Theta}^{(N_C)}\right\}$
  $\boldsymbol{\Xi} \leftarrow \left\{\boldsymbol{\Theta}^{(1)},\ldots,\boldsymbol{\Theta}^{(N_C)}\right\}$
  `LP ← ParallelPricingKernel<<<`N_c, N_b`>>>`$\left(\boldsymbol{\Xi}, \{(F,K,T)_i\}_{i=1}^{I}\right)$
  for $g = 1,\ldots,G$ **do**
    for $c = 1,\ldots,N_C$ **do**
      `Randomly choose` $j,k,m \in \{1,\ldots,N_A\}$ `s.t.` j,k,m,i `are all distinct`
      $\boldsymbol{\Theta}^{(n)} \leftarrow \boldsymbol{\Theta}^{(j)} + F \times (\boldsymbol{\Theta}^{(k)} - \boldsymbol{\Theta}^{(m)})$
      for $j = 1,\ldots,J$ **do**
        `Generate` $a_j^i \sim U[0,1]$
        if $a_j^i < CR$ **then**
          $\hat{\Theta}_j^{(c)} \leftarrow \Theta_j^{(n)}$
        else
          $\hat{\Theta}_j^{(c)} \leftarrow \Theta_j^{(c)}$
        end if
      end for
      $\hat{\boldsymbol{\Xi}}_c \leftarrow \hat{\boldsymbol{\Theta}}^{(c)}$
    end for
    `LN ←ParallelPricingKernel<<<`N_c, N_b`>>>`$\left(\hat{\boldsymbol{\Xi}}, \{(F,K,T)_i\}_{i=1}^{I}\right)$
    for $c = 1,\ldots,N_C$ **do**
      if `LN[c] < LP[c]` **then**
        `LC[c] ← LN[c]`
        $\boldsymbol{\Xi}_c \leftarrow \hat{\boldsymbol{\Xi}}_c$
      end if
    end for
  end for
  return $\boldsymbol{\Theta}^{(c)} \in \boldsymbol{\Xi} \mid L\left(\boldsymbol{\Theta}^{(c)}\right) \leq L\left(\boldsymbol{\Theta}^{(c')}\right), c' = 1,\ldots,N_C$
end procedure

4.4 Numerical Data And Results

We now give the results from our numerical tests on the model proposed in section 4.2, along with a detailed description of the data used and the implementation details of the computational method described in section 4.3.

4.4.1 Technical Implementation Details

The code for the numerical tests was written in C++ in Visual Studio 2017 for execution on the host system and CUDA-C for execution on the GPU with version 9 of CUDA. The machine used was a 64 bit Lenovo ThinkPad T470p with a i7-780HQ dual core at 2.9GHz and 16.0GB RAM. The graphics card used was NVIDIA GeForce 940MX.

For obtaining an approximate global minimum of the least squares objective function (4.13) we used the differential evolution procedure given in Algorithm 4 with 50 candidate solutions that were initially randomly generated and a termination criterion of 100 iterations. The crossover probability was set at 0.9 and the weight parameter at 0.8.

For refining the best solution from the DE procedure we used box-constrained Levenberg-Marquardt routine from the ALGLIB library with a central finite difference approximation to the gradient. The gradient tolerance was set at 2/3 of the 64 bit machine precision. For the Fourier inversion, we used the Gauss-Legendre quadrature with 128 nodes.

The Differential Evolution routine and the Levenberg-Marquardt routine shared the same parameter search constraints, which are given in Table 4.1.

	min	max
κ	2.0	50.0
η	0.01	0.90
ν	0.01	0.90
σ	5.0	30.0
ρ	-0.99	0.99
λ	0.0	20.0
τ	0.0	1.0

Table 4.1: Variable Constraints. These represent the intervals within which we constrained the optimization procedure for each parameter.

As of yet, the library of mathematical functions in CUDA is rather limited. We therefore had to implement complex valued versions of the Gaussian error function, the Gamma function and the confluent hypergeometric function explicitly in order to be able to compute the characteristic function of the seasonal 3/2 model on the GPU. For the implementation of the Gaussian error function we ported code from the Faddeeva module made available by the AB-Initio physics research group at MIT[22], and for the implementation of the gamma function and confluent hypergeometric function we used the respective formulas given in Abramowitz and Stegun [1964].

4.4.2 Data

We compared the empirical fit of the homogeneous 3/2 model defined by (4.1) and the seasonal 3/2 model defined by (4.2) to market prices using weekly (first trading day of the week) market prices of options written on corn and wheat futures for the period from 01/04/2016 to 10/16/2017. From this data set, we retained only out-of-the-money (OTM) call and put options for which there was strictly positive open interest, and the settlement price was at least 0.2.

Furthermore, we grouped the option contracts from each trading day according to the expiration date of the underlying futures contract, and categorized these as 'short' (with futures expiration date less than 100 days from trading day), 'medium' (with futures expiration date between 100 and 200 days from trading day), and 'long' (with futures expiration date between 200 and 365 days from trading day).

Within each of these expiration categories for a given trading day we chose the option maturities for which at least 5 OTM calls and 5 OTM puts satisfied the liquidity criteria listed above. Trading days that did not contain enough liquid option such that the aforementioned procedure yielded at least one representative maturity for each expiration category were dropped from the sample. This left us with 91 trading days for the corn futures options and 94 trading days for the wheat futures options. Out of these, we chose the 5 puts and 5 calls which were closest to the money for the shortest maturity, and the 5 puts and 5 calls for each of the longer maturities were selected (to the extent possible given the occasionally limited number of liquid options) in such a way that the aggregate price was roughly the same across the different maturities. This was done in an effort to prevent the more expensive long dated options from disproportionately influencing the least squares objective function.

22. See `http://ab-initio.mit.edu/wiki/index.php/Faddeeva_Package`

4.4.3 European Option Price Approximation

While the pricing formulas given in the preceding section assume the options are European, the options in our market data are American options. For the call options these option types coincide in price, whereas for the put options they differ with non-zero interest rates. To address this discrepancy, we follow the work of Trolle and Schwartz [2009] and Arismendi et al. [2016] and use the inverted formula of Barone-Adesi-Whaley (BAW) (Barone-Adesi and Whaley [1987]) to compute the prices of the corresponding European put options implied by the American put options in our data set.

We begin by recalling[23] the pricing PDE of the Black-76 model (Black [1976]),

$$\frac{\partial J}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 J}{\partial S^2} - rJ = 0. \tag{4.15}$$

Now we define the following:

$$\begin{aligned} \gamma &= \frac{1}{2}\left(1 - \sqrt{1 + \frac{8r}{\left(1 - e^{-rT}\right)\sigma^2}}\right), \\ A &= -\frac{S^*}{\gamma}\left(1 - e^{-rT}\mathrm{N}[-d_1(S^*)]\right), \\ d_1(S) &= \frac{\ln(S/K) + \sigma^2 T/2}{\sigma\sqrt{T}}. \end{aligned} \tag{4.16}$$

Then the BAW approximation for the American put option price π^{AP} given the corresponding European put option price π^{EP} is given by

$$\pi^{AP}(S_0, T, K) = \begin{cases} \pi^{EP}(S_0, T, K) + A\left(\frac{S_0}{S^*}\right)^{\gamma} & S_0 > S^*, \\ K - S_0 & S_0 \leq S^*. \end{cases} \tag{4.17}$$

The variable S^* is the stock price below which the option should be exercised. A closed form solution for S^* is not available, however, we can compute it by numerically solving

$$K - S^* = \pi^{EP}(S^*, T, K) - \left(1 - e^{-rT}\mathrm{N}[-d_1(S^*)]\right)\frac{S^*}{\gamma} \tag{4.18}$$

23. The BAW approximation is derived for the Black-Scholes model, and since we are working with futures as the underlying asset with zero cost-of-carry, the Black-Scholes pricing PDE is simplified to that of the Black-76 model. Note that the application of this formula to the prices in our data sample does not enforce the assumptions of the Black-Scholes model since the implied volatility is not constrained to be constant across the different option prices we are computing.

using a root search algorithm. Once we have computed S^* we can then derive $\pi^{EP}(S_0, T, K)$ from (4.17) since we know π^{AP} from the data.

We note that while the use of the BAW formula for retrieving the price of a European option from a corresponding American option is used throughout the previously mentioned literature on futures options pricing, other methods could be used as well. In particular, lattice methods, when applicable, are known to be exceptionally fast for computing the prices of American options and could be implemented in a similar fashion here to obtain π^{EP}. See e.g. Burkovska et al. [2018] for an exploration of this methodology.

4.4.4 Results

For each trading day, both the homogeneous 3/2 model and the seasonal 3/2 model were calibrated against the set of benchmark options obtained from the data filtering procedure described above, and the root mean squared price (RMSP) error and the root mean squared implied volatility (RMSIV) error recorded. The summary statistics for these two error measurements are given in Tables 4.4 and 4.5 for the corn options and wheat options, respectively. Furthermore, the estimated parameter values obtained from each calibration run were recorded. The summary statistics for these values are given in Tables 4.2 and 4.3 for the corn options and wheat options, respectively.

As can be seen from Tables 4.4 and 4.5, the seasonal 3/2 model outperforms the homogeneous 3/2 model across all maturity groups and moneyness categories both in terms of both error measures for both the corn futures options and the wheat futures options. For the corn futures options the seasonal 3/2 model gives roughly 32% lower RMSP error and 31% RMSIV error compared to the homogeneous 3/2 model. And for the wheat futures options the RMSP and RMSIV errors are approximately 62% and 35% lower, respectively, for the seasonal 3/2 model.

Model	Statistic	κ	η	v_0	σ	ρ	λ	τ
H32	mean	6.180	0.633	0.060	11.399	0.467	*	*
	std	6.489	0.325	0.047	5.335	0.091	*	*
	max	50.000	0.891	0.314	30.000	0.816	*	*
	min	2.000	0.010	0.016	5.000	0.248	*	*
S32	mean	15.811	0.182	0.105	14.965	0.587	6.179	0.189
	std	18.216	0.303	0.161	7.297	0.143	7.483	0.083
	max	50.000	0.773	0.824	29.638	0.899	20.000	0.388
	min	2.000	0.001	0.009	5.000	0.357	0.000	0.000

Table 4.2: Parameter summary for the calibration of the homogeneous 3/2 model (H32) and the seasonal 3/2 model (S32) using options on corn futures.

Model	Statistic	κ	η	v_0	σ	ρ	λ	τ
H32	mean	7.888	0.656	0.070	11.154	0.477	*	*
	std	12.695	0.280	0.064	3.079	0.055	*	*
	max	50.000	0.900	0.416	19.476	0.581	*	*
	min	2.000	0.062	0.023	8.200	0.385	*	*
S32	mean	9.490	0.381	0.089	15.708	0.544	4.441	0.196
	std	10.698	0.399	0.053	3.243	0.077	4.295	0.118
	max	50.000	0.900	0.283	23.462	0.728	16.316	0.491
	min	2.119	0.001	0.024	9.807	0.330	0.000	0.000

Table 4.3: Parameter summary for the calibration of the homogeneous 3/2 model (H32) and the seasonal 3/2 model (S32) using options on wheat futures.

Model	Strike	Short	Medium	Long	All
H32	K<F	0.688, 0.022	0.730, 0.019	0.697, 0.011	0.707, 0.016
	K>F	0.861, 0.032	0.648, 0.015	0.593, 0.011	0.655, 0.017
	All	0.779, 0.028	0.690, 0.017	0.647, 0.011	0.682, 0.016
S32	K<F	0.220, 0.010	0.395, 0.011	0.440, 0.007	0.401, 0.009
	K>F	0.273, 0.014	0.462, 0.012	0.588, 0.011	0.514, 0.012
	All	0.247, 0.012	0.430, 0.012	0.519, 0.009	0.461, 0.011

Table 4.4: Pricing errors for the homogeneous 3/2 model (H32) and the seasonal 3/2 model (S32) for options on corn futures. The first entry in each column is the root mean squared price error, while the second entry in each column is the root mean squared implied volatility error.

Model	Strike	Short	Medium	Long	All
H32	K<F	0.771, 0.019	0.602, 0.014	0.792, 0.009	0.725, 0.013
	K>F	0.967, 0.025	0.654, 0.013	0.949, 0.010	0.857, 0.014
	All	0.875, 0.023	0.629, 0.013	0.874, 0.009	0.794, 0.014
S32	K<F	0.234, 0.011	0.275, 0.008	0.361, 0.005	0.313, 0.007
	K>F	0.215, 0.020	0.261, 0.006	0.315, 0.005	0.281, 0.010
	All	0.225, 0.016	0.268, 0.007	0.339, 0.005	0.298, 0.009

Table 4.5: Pricing errors for the homogeneous 3/2 model (H32) and the seasonal 3/2 model (S32) for options on wheat futures. The first entry in each column is the root mean squared price error, while the second entry in each column is the root mean squared implied volatility error.

4.5 Conclusion

In this chapter we proposed a modified 3/2 stochastic volatility model for assets that exhibit seasonal trends in volatility, such as futures on corn and wheat. To our knowledge, this is the first non-affine stochastic volatility model proposed in the literature that incorporates deterministic seasonal trends in the volatility process. We also derived the characteristic function for this model in closed form, which can be used to greatly reduce the computational load of using the model for pricing purposes.

We studied the in-sample empirical fit of this new model, as well as the original 3/2 model, to market data consisting of options written on corn and wheat futures. We implemented a robust minimization procedure for calibrating the models that mixed heuristic global minimum search with a Quasi-Newton method. The most computationally demanding aspects of the code were written in CUDA-C++ and executed on a GPU to prevent the procedure from becoming infeasibly slow.

Our tests revealed that the seasonal 3/2 model significantly outperformed the homogeneous 3/2 model across all strikes and maturity groups, both in terms of root-mean-squared error as well as mean implied volatility error.

This is a promising result, particularly since the new model contains no additional stochastic factors beyond the original 3/2 model, which means that no additional computational cost is incurred for derivative pricing that requires more general numerical approaches such as finite difference methods or Monte Carlo simulations. The characteristic function we derived for the model is mathematically tractable, and by using standard analytic approximations to the Gaussian error function that is embedded in it we avoid any numerical integration beyond the Fourier inversion itself.

Chapter 5

A Generalized Weighted Monte Carlo Calibration Method

5.1 Introduction

So far, we have focused on the calibration of the 3/2 stochastic volatility model and its more general seasonal modification primarily in terms of the characteristic function. However, not all models come with a known characteristic function in closed form, which means we need to resort to different computational methods for the calibration process.

If the model prices can be computed using a Monte Carlo simulation the weighted Monte Carlo (WMC) method by Avellaneda et al. [2001] gives us a way to calibrate the model in such a way that we obtain from that model a set of weighted paths which can reproduce exactly (or almost exactly as in a least squares approach) the prices of the derivatives used in the model calibration.

More specifically, the method consists of simulating a set of price paths using the no-arbitrage model that is to be calibrated, and calculating a new (risk neutral) probability measure for this set of paths that reproduces the observed market prices of benchmark options exactly, or almost exactly as in the case of a least squares approach. As we tend to have more paths than benchmark options in a simulation, such a measure is not uniquely defined in general. This problem is solved in the original WMC method by selecting the measure closest to the (uniform) prior measure in terms of relative entropy (Kullback and Leibler [1951]), also referred to as Kullback-Leibler divergence.

While relative entropy is a well established concept within information theory, it does not by itself offer any particular economic intuition in the context of the WMC method. The implication of this is that there is no obvious way to discriminate between relative entropy and any other kind of divergence when choosing how to weight the sample paths in a manner that is economically justified. However, as we discuss in greater detail in the following sections, the entropy minimization problem of the WMC method is mathematically equivalent to a portfolio choice problem for an investor with expected exponential utility. And by considering the utility version of the WMC, we open up the possibility of exploiting the theoretically and empirically mature field of consumption based asset pricing (Ludvigson [2011]; Carmona [2009]) to develop refinements of the WMC method.

As we explain in greater detail in the following sections, the WMC calibration method can in theory reweight the paths simulated from the model to be calibrated in such a way that they can reproduce the market prices of the benchmark options used in the calibration procedure exactly. However, as we observe in our numerical tests for two popular option pricing models (the Black Scholes model and the stochastic volatility model of Heston), the accuracy drops quite quickly for out-of-sample options as we move away from the strike range and maturity range of the benchmark options.

The contribution of this chapter consists of formulating a more general version of the WMC which in our numerical tests produces a far better fit to the whole range of options available on the underlying asset than the original WMC. It achieves this by first splitting the paths into segments by the maturities present in the set of benchmark options, and then applying a probability distortion transformation to the prior distribution associated with these path segments. This probability distortion transformation is inspired by the work of Kahneman and Tversky [1992] on Cumulative Prospect Theory (CPT), but differs in that our approach preserves the additivity of the probability measure, whereas the probability distortion in the CPT model leads to a non-additive one which results in certain tractability challenges.

The remainder of this chapter is as follows: in section 5.2 we give a general formulation of the weighted Monte Carlo method, in section 5.3 we discuss the relationship between entropy minimization and utility maximization and give an alternative formulation of the entropy minimization problem as a portfolio choice problem, in section 5.4 we propose a weighted Monte Carlo method that incorporates multiple weights-per-path and rare event probability distortion, and in section 5.5 we give the numerical results for the different methods.

5.2 An Overview of the Weighted Monte Carlo Method

We begin by briefly describing the weighted Monte Carlo method in general terms. Our discussion follows the same reasoning as that presented in a more comprehensive format in Avellaneda et al. [2001]), with the exception that we do not assume that the prior distribution is uniform. So the original WMC method is a special case of what is presented in this section.

Given a filtered probability space $(\Omega, \mathscr{F}, \boldsymbol{F}, \mathbb{P})$, a finite time horizon T, and an adapted price process $S = \{S_t\}_{t\in[0,T]}$ we can view the set of N paths produced by a Monte Carlo simulation of S_t as a discrete approximation to the distribution of S at time t. The general idea behind the weighted Monte Carlo approach is to reweight the sampled paths in such a way that the new distribution is as statistically close as possible to the original one, while at the same time reproducing the observed market prices at time 0 for a set of contingent claims on S. Here, the notion of statistical distance is taken to be the Kullback-Leibler divergence[24], which for two discrete probability measures, $\mathbb{P}$ and $\mathbb{Q}$ is given by

$$D_{KL}(\mathbb{Q}|\mathbb{P}) = \sum_i q_i \ln\left(\frac{q_i}{p_i}\right).$$

A no-arbitrage model described at each $t \in [0,T]$ by S_t which we want to calibrate using the weighted Monte Carlo method is referred to hereafter as the initial model. Assume we have a set of K benchmark options[25] on S we want to use to reweight the simulated paths of the model. We begin by simulating N paths from S and calculating the payoff matrix

$$\boldsymbol{G} = \begin{pmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,K} \\ g_{2,1} & g_{2,2} & \cdots & g_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N,1} & g_{N,2} & \cdots & g_{N,K} \end{pmatrix}, \tag{5.1}$$

where g_{ik} is the payoff from option k when path i is realized.

Assuming the simulated paths are initially assigned prior weights given by $\boldsymbol{p} = (p_1, \ldots, p_N)$ then the new weights $\boldsymbol{q} = (q_1, \ldots, q_N)$ in the exact-fit version

24. However, other types of divergences have been studied, see Friedman et al. [2013].
25. Any type of option will do as long as its payoff along a given sample path is completely determined by that path. This excludes, for example, American style options.

are calculated as the solution to

$$\begin{aligned} &\min_{\boldsymbol{q}} D(\boldsymbol{q}|\boldsymbol{p}) = \min_{\boldsymbol{q}} \left\{ \sum_{i=1}^{N} q_i \ln\left(\frac{q_i}{p_i}\right) \right\} \\ &\text{s.t. } \pi_k = \sum_{i=1}^{N} q_i g_{ik},\ k = 1,\ldots,K \end{aligned} \tag{5.2}$$

where π_k is the price of benchmark option k. Note that throughout our discussion we will use $\mathbb{Q}$ and $\boldsymbol{q}$ (and $\mathbb{P}$ and $\boldsymbol{p}$) interchangeably but write the former when we want to emphasize it as a probability measure and the latter when we want to emphasize it as a set of path weights

Using the Lagrange multiplier approach, this can be rewritten as the dual problem

$$\min_{\lambda} \max_{\boldsymbol{q}} \left\{ -D(\boldsymbol{q}|\boldsymbol{p}) + \sum_{k=1}^{K} \lambda_k \left(\sum_{i=1}^{N} q_i g_{ik} - \pi_k \right) \right\}. \tag{5.3}$$

Looking at the first order conditions for the inner problem we see that a solution is given by

$$q_i = \frac{p_i}{Z(\lambda)} \exp\left(\sum_{k=1}^{K} g_{ik} \lambda_k \right), \tag{5.4}$$

where

$$Z(\lambda) = \sum_{m=1}^{N} p_m \exp\left(\sum_{k=1}^{K} g_{mk} \lambda_k \right),$$

is a normalization factor to ensure the distribution sums to 1. Using (5.4) we can simplify (5.3) to get

$$\min_{\lambda} \left\{ \ln\left(Z(\lambda)\right) - \sum_{k=1}^{K} \pi_k \lambda_k \right\}. \tag{5.5}$$

This problem requires numerical methods to solve. As the objective function is smooth and convex, the method of choice is a gradient-based minimization procedure such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Broyden [1970]) which requires the computation of the partial derivatives of (5.5). They are given by

$$\frac{\partial W}{\partial \lambda_j} = \frac{1}{Z(\lambda)} \sum_{i=1}^{N} g_{ij} p_i \exp\left(\sum_{k=1}^{K} g_{ik} \lambda_k \right) - \pi_j, \tag{5.6}$$

for $j = 1,\ldots,K$.

Finally, although a unique solution should always exist for (5.2) given an arbitrage free market, the presence of asynchronous and noisy data can lead to

problems for the optimization procedure, with one instrument being bought or sold in quantities much larger than the rest. A remedy to this problem is adding a regularization term to the objective function. The term we use in our tests is a simple quadratic term,

$$\chi^2_w = \frac{1}{2}\sum_{k=1}^{N}\frac{1}{w_k}\lambda_k^2, \tag{5.7}$$

where $w_1,\ldots,w_K$ are penalization weights that determine the influence of option k on the calibration. The resulting optimization problem in (5.5) is given by

$$\min_{\lambda}\left\{\ln(Z(\lambda)) - \sum_{k=1}^{K}\pi_k\lambda_k + \chi^2_w\right\}. \tag{5.8}$$

5.3 The Weighted Monte Carlo Method as a Utility Maximization Problem

Using the observed price information of assets such as stocks and options in a financial market to derive (or approximate) the stochastic discount factor of that market is a central task in the field of asset pricing (for a standard reference, see e.g. Cochrane [2000]). Within the field of financial economics, this is commonly done in the setting of consumption based pricing, where the market is treated like a single representative investor with a given utility function over uncertain future cash flows. The stochastic discount factor in this setting is proportional to the marginal utility of the representative investor over her optimal consumption in each possible future state. Given a set of contingent claims in the market, this optimal consumption process can be calculated by solving for the optimal portfolio of the investor.

In more concrete terms, consider the following one period[26] model. We assume the investor maximizes expected utility, with investment horizon T. The assets available to the investor form an arbitrage-free market of total initial wealth W_0 with an associated probability space $(\Omega,\mathscr{F},\mathbb{P})$. Let the investor's utility function over deterministic outcomes realized at time T be given by a continuously differentiable function $u:\mathbb{R}\to\mathbb{R}$ such that $u'(x)>0$ and $u''(x)<0$ for all $x\in\mathbb{R}$, and let $\mathscr{A}(W_0)$ denote the set of all contingent claims in the market which can be financed with initial wealth W_0. If we denote with X_T^* the optimal solution to the portfolio selection problem

$$\sup_{X_T\in\mathscr{A}(W_0)}\mathbb{E}^{\mathbb{P}}[u(X_T)], \tag{5.9}$$

26. Here, 'one period' refers to the investor only choosing a portfolio at time 0. This does not restrict the assets from evolving continuously from time 0 to time T with realized payoffs in between.

there exists $\beta \in \mathbb{R}^+$ such that for every contingent claim X_T in the market we have that the market price of X_T is given by $\pi_X = \mathbb{E}^{\mathbb{P}}[\beta u'(X_T^*)X_T]$. We can write the expectation in more compact form through a change of measure using the Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}} = u'(X_T^*)$ to obtain $\pi_X = \mathbb{E}^{\mathbb{Q}}[\beta X_T]$.

In mathematical terms, the correspondence between this type of utility maximization and entropy minimization is well known (see e.g. Nau et al. [2009]). It is therefore reasonable to ask precisely how this relationship enters in the weighted Monte Carlo setting. As is explained in Avellaneda et al. [2001] the Arrow-Debreu prices that correspond to the measure computed from (5.2) coincide with the marginal utilities for consumption obtained by solving (5.9) when u is the exponential utility function.

More specifically, solving for the Lagrange multipliers λ_k in (5.3) is equivalent to finding the optimal portfolio weights for a utility maximizer with exponential utility (Samperi [1997]), where we can think of the benchmark instruments as the assets in the market, and the set of sample paths generated by the initial model as the statespace of the market. The relationship between the two is more precisely as follows: the optimal portfolio weights given by $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ for an investor with a utility function given by

$$U(\boldsymbol{\phi}) = -\sum_{i=1}^{N} p_i \exp\left(-\sum_{k=1}^{K} \phi_k (g_{ik} - \pi_k)\right) \tag{5.10}$$

are related to the optimal Lagrange multipliers $\lambda_1, \ldots, \lambda_K$ by $\phi_k = -\lambda_k$ for $k = 1, \ldots, K$. Consequently, if we assume that the utility maximizer is a representative investor, then the path weighting through entropy minimization is mathematically equivalent to deriving the stochastic discount factor in this market through solving for the optimal portfolio of the representative investor.

Before elaborating further on how the weighted Monte Carlo method fits with the standard consumption-based asset pricing approach, we need to introduce some additional terminology. Given an underlying statespace Ω, and a σ-algebra $\mathscr{F}$ defined on Ω, we refer to the probability measure that describes the true probability of the different events in $\mathscr{F}$ as the *objective* measure. In contrast, we refer to any probability measure that is absolutely continuous with respect to the objective measure, but absorbs to some extent risk premiums that exist in the market, as a *subjective* measure. An example of a subjective measure would be a risk neutral measure, i.e., a probability measure $\mathbb{Q}$ such that the market price of any contingent claim in the market is the discounted expected value of the claim with respect to $\mathbb{Q}$. Lastly, we use "pricing measure" and "risk neutral measure" interchangeably.

The standard way to formulate a consumption-based asset pricing model is to assume that the investor observes the objective measure $\mathbb{P}$ associated with the statespace. This means the expected utility is calculated under the $\mathbb{P}$ measure. On the other hand, option pricing models are typically calibrated against existing option contracts, making the probability measure corresponding to the sample paths of the initial model subjective, as opposed to objective.

However, this apparent discrepancy is dissolved with the realization that we are effectively deriving a "decomposed" stochastic discount factor. To clarify, let $\mathbb{L}$ be the subjective probability measure corresponding to the uniform weighting of the Monte Carlo simulation paths generated by the initial model (i.e., R is the subjective measure under which the sample path approximation to the initial model is given). Let $\mathbb{P}$ denote the objective probability measure corresponding to these paths. If our model is arbitrage free, we expect that $\mathbb{L} \ll \mathbb{P}$, i.e., that $\mathbb{L}$ is absolutely continuous with respect to $\mathbb{P}$. Next, let $\mathbb{Q}$ be the subjective probability measure corresponding to the weighting of the sample paths computed by the WMC method. Again, the absence of arbitrage means we must have $\mathbb{Q} \ll \mathbb{L}$. The Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{L}}$ is proportional to the stochastic discount factor $u'(X_T^*)$ we obtain from the utility maximization equivalent of the WMC entropy minimization. Given that $\mathbb{Q} \ll \mathbb{L} \ll \mathbb{P}$, then by the measure-theoretic chain rule, we have that

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{d\mathbb{Q}}{d\mathbb{L}} \frac{d\mathbb{L}}{d\mathbb{P}} \text{ - } \mathbb{P} \text{ a.s.}$$

In other words, changing the measure from $\mathbb{P}$ to $\mathbb{L}$ and then from $\mathbb{L}$ to $\mathbb{Q}$ is the same (a.s.) as changing it straight from $\mathbb{P}$ to $\mathbb{Q}$. So whether we derive the risk neutral pricing distribution $\mathbb{Q}$ straight from $\mathbb{P}$ as in the standard representative investor pricing approach, or by first deriving $\mathbb{L}$ from $\mathbb{P}$ and then $\mathbb{Q}$ from $\mathbb{L}$ as in the WMC approach, we end up with the same results.

With the theoretical justification out of the way, we now give the utility maximization equivalent of the formulation presented in section 5.2. Let $u(\cdot)$ be defined as in (5.9). The statespace Ω consists of the paths we simulate from the initial model, and $\mathbb{P}$ is whatever weighting we attach to the paths to represent the prior measure. The market consists of the benchmark instruments we use for the calibration. We denote with $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ the vector of prices for the benchmark instruments, and with $\boldsymbol{G} = (G_1, \ldots, G_K)$ the vector of their stochastic payoffs. With this in mind, we formulate the portfolio selection problem for the investor as

$$\max_{\boldsymbol{\phi}} \mathbb{E}^{\mathbb{P}} \left[u(\boldsymbol{\phi}^\top \boldsymbol{G}) \right] \quad \text{s.t.} \quad \boldsymbol{\phi}^\top \boldsymbol{\pi} \leq W_0, \tag{5.11}$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_K)$ is the portfolio choice. Since an unconstrained optimization problem generally allows more efficient computational methods, and the budget constraint can be assumed to hold with equality, we can rewrite (5.11) as

$$\max_{\boldsymbol{\phi}} \sum_{i=1}^{N} p_i u\left(\sum_{k=2}^{K} \phi_k \left(g_{ik} - g_{i1}\frac{\pi_k}{\pi_1}\right) + \frac{g_{i1}W_0}{\pi_1}\right), \tag{5.12}$$

where we have substituted

$$\phi_1 = \frac{1}{\pi_1}\left(W_0 - \sum_{k=2}^{K} \phi_k \pi_k\right)$$

in the maximization problem above, and g_{ik} denotes the payoff from option k when path i is realized. The first order conditions are given by

$$\sum_{i=1}^{N}\left[p_i\left(g_{ih} - g_{i1}\frac{\pi_h}{\pi_1}\right)u'\left(\sum_{k=2}^{K} \phi_k \left(g_{ik} - g_{i1}\frac{\pi_k}{\pi_1}\right) + \frac{g_{i1}W_0}{\pi_1}\right)\right] = 0, \tag{5.13}$$

for $h = 1, \ldots, K$. The solution $\boldsymbol{\phi}^* = (\phi_1^*, \ldots, \phi_K^*)$ to this problem, which needs to be computed using numerical methods, is the optimal portfolio choice for the representative investor with initial wealth W_0.

The pricing kernel is now obtained by plugging $\left(\boldsymbol{\phi}^*\right)^\top \boldsymbol{G}$ into $u'(\cdot)$, and the corresponding change of measure gives us the new risk neutral distribution we are after. More precisely, the new weights q_i for $i = 1, \ldots, N$ are given by

$$q_i = \frac{p_i u'\left(\sum_{k=2}^{K} \phi_k^* \left(g_{ik} - g_{i1}\frac{\pi_k}{\pi_1}\right) + \frac{g_{i1}W_0}{\pi_1}\right)}{\sum_{i=1}^{N} p_i g_{i1}}\pi_1. \tag{5.14}$$

The least squares setup is the following:

$$\max_{\boldsymbol{\phi}} \sum_{i=1}^{N} p_i u\left(\sum_{k=2}^{K} \phi_k \left(g_{ik} - g_{i1}\frac{\pi_k}{\pi_1}\right) + \frac{g_{i1}W_0}{\pi_1}\right) - \frac{1}{2}\sum_{k=2}^{K} w_k \phi_k^2 - \frac{1}{2}\left(\frac{1}{\pi_1}\left(W_0 - \sum_{k=2}^{K} \phi_k \pi_k\right)\right)^2, \tag{5.15}$$

with the first order conditions given by

$$\sum_{i=1}^{N}\left[p_i\left(g_{ih} - g_{i1}\frac{\pi_h}{\pi_1}\right)u'\left(\sum_{k=2}^{K} \phi_k \left(g_{ik} - g_{i1}\frac{\pi_k}{\pi_1}\right) + \frac{g_{i1}W_0}{\pi_1}\right)\right] - w_h\phi_h + \frac{\pi_h}{\pi_1}\left(W_0 - \sum_{k=2}^{K} \phi_k \pi_k\right) = 0, \tag{5.16}$$

with $h = 1,\ldots,K$ and the weights q_i for $i = 1,\ldots,N$ given by (5.14) as before.

As previously mentioned, if we set $u(x) = -e^{-x}$ and $p_i = \frac{1}{N}$ for $i = 1,\ldots,N$, the formulation is mathematically equivalent to the original weighted Monte Carlo method of Avellaneda et al. [2001], where the measure of statistical distance is given by the Kullback-Leibler divergence and the prior measure is uniform. The initial wealth W_0 in this case does not affect the solution, since $u(x)$ is translation invariant. In the general case, a straightforward choice for W_0 that conforms to the representative investor model is the "market wealth", i.e., the value of the underlying assets.

5.4 Calibration with Probability Distortion

As can be seen from Figures 5.3 and 5.4 in section 5.5, the implied volatility we obtain from the original weighted Monte Carlo method turns out to be much lower for the far-OTM options than what is implied by the market prices. From the consumption-based asset pricing perspective this could in theory be explained by positing that the preferences represented by (5.10), and by extension of duality the relative entropy formulation in (5.2) are not risk averse enough. More specifically, (5.10) trivially includes a coefficient of risk aversion equal to one. Adding an explicit coefficient of risk aversion to the formulation in (5.10) and then deriving the corresponding generalized version of (5.5) using the Legendre transform then gives us a divergence measurement that contains a parameter which directly affects the tail thickness of the derived subjective probability measure represented by $\boldsymbol{q}$, with higher values for the risk aversion coefficient translating to thicker tails.

Preliminary numerical tests revealed, however, that the risk aversion coefficient by itself had barely a noticeable effect on the tail thickness for the range of values which still allowed a decent fit with the benchmark instruments.

From a decision-theoretic perspective we can, however, accommodate the thick tails implied by the market prices by positing that the market overweights the probability of large market movements with respect to the initial model. To implement this idea, we apply a probability distortion that is partly similar to that introduced by Kahneman and Tversky [1992] in their work on Cumulative Prospect Theory. This generalization turns out to give vastly better empirical performance than the original method as well as its purely utility function-based modifications in our tests.

As compared to expected utility, two novel features characterize CPT preferences. The first feature is a utility function over deterministic outcomes that is concave over gains and convex over losses with respect to a given

reference point of wealth. The second feature is a distortion function that is applied to the cumulative distribution function of the physical measure, yielding a non-additive measure. The general CPT specification is given by

$$U(X) = \int_0^{+\infty} W^+(\mathbb{P}(u^+(X) > x))dx - \int_0^{+\infty} W^-(\mathbb{P}(u^-(X) > x))dx, \quad (5.17)$$

where W^+, W^-, u^+ and u^- are the probability distortion functions and utility functions over deterministic outcomes for the gain and loss domains, respectively, and $\mathbb{P}$ is a probability measure. The distortion function we chose to implement for the numerical tests is the one introduced by Prelec [1998], which gave considerably better results than the original distortion function by Kahneman and Tversky, and is given by

$$v^+(P) = \exp\left\{-\gamma^+(-\ln(P))^{\delta+}\right\} \quad (5.18)$$

for the gains domain, and

$$v^-(P) = \exp\left\{-\gamma^-(-\ln(P))^{\delta-}\right\} \quad (5.19)$$

for the loss domain. Here, δ^+ and δ^- correspond to the curvature of the distortion, while γ^+ and γ^- correspond to the elevation of the distortion, with a value of 1.0 for all parameters corresponding to a non-distorted probability measure, as demonstrated in Figures 5.1 and 5.2. See Prelec [1998] for a comprehensive discussion on the interpretation of these parameters.

We can adapt the formulation in (5.17) to a discrete statespace in the following way. Let $\Omega = \{x_{-m}, \ldots, x_0, \ldots, x_n\}$ be the set of possible outcomes, such that $x_{-m} \leq x_{-m+1} \leq \ldots \leq x_{n-1} \leq x_n$ with $x_0 = 0$ by convention, and let p_i be the probability of outcome x_i for $i = -m, \ldots, n$. Furthermore, let $u^- : \Omega^- \to \mathbb{R}$, be a strictly increasing convex function, with $\Omega^- = \{x_{-m}, \ldots, x_{-1}\}$, and $u^+ : \Omega^+ \to \mathbb{R}$, be a strictly increasing concave function, with $\Omega^+ = \{x_0, \ldots, x_n\}$. Furthermore, let $v^- : [0,1] \to [0,1]$ and $v^+ : [0,1] \to [0,1]$ be two strictly increasing functions such that $v^-(0) = v^+(0) = 0$ and $v^-(1) = v^+(1) = 1$ If we denote by $X = (x_{-m}, p_{-m}; \ldots; x_{-1}, p_{-1}; x_0, p_0; x_1, p_1; \ldots; x_n, p_n)$ the prospect X, then the Cumulative Prospect value of X is given by

$$U(X) = \sum_{i=-m}^{n} v_i u(x_i), \quad (5.20)$$

where

$$u(x_i) = \begin{cases} u^+(x_i), & \text{if } 0 \leq i \leq n, \\ u^-(x_i), & \text{if } -m \leq i < 0, \end{cases} \quad (5.21)$$

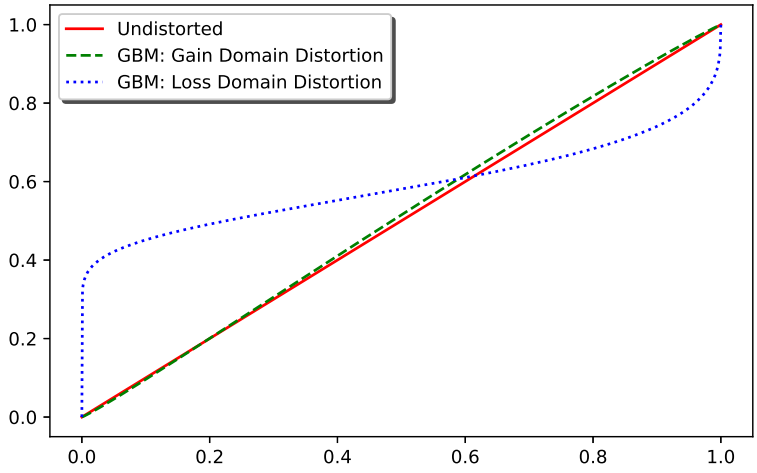

Figure 5.1: Mean probability distortion values for the gain domain ($\gamma = 1.051, \delta = 0.97$) and the loss domain ($\gamma = 0.316, \delta = 0.611$) for the GBM model as presented in Table 5.2.

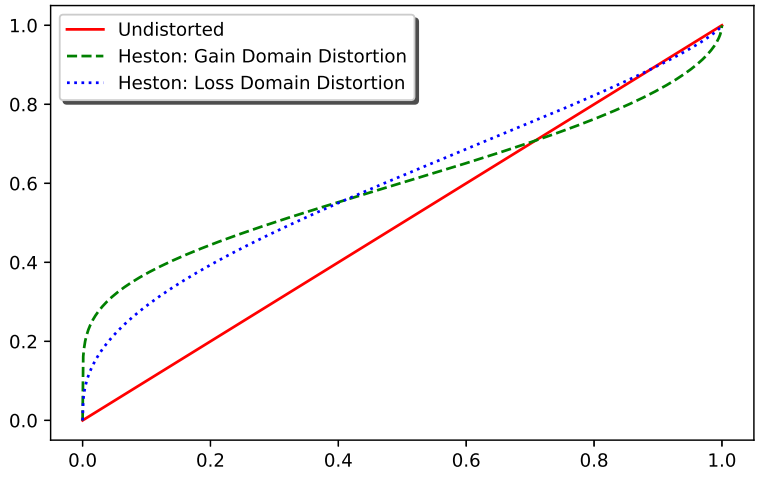

Figure 5.2: Mean probability distortion values for the gain domain ($\gamma = 0.555, \delta = 0.623$) and the loss domain ($\gamma = 0.791, \delta = 0.640$) for the Heston model as presented in Table 5.2.

and

$$w_i = \begin{cases} w^+(p_i + \ldots + p_n) - w^+(p_{i+1} + \ldots + p_n), & \text{if } 0 \le i \le n, \\ w^-(p_{-m} + \ldots + p_i) - w^-(p_{-m} + \ldots + p_{i-1}), & \text{if } -m \le i < 0. \end{cases} \tag{5.22}$$

On a theoretical level, CPT has been shown to produce potential resolutions to a wide array of empirical puzzles in finance and economics (Camerer [2004]; Barberis [2013]; Fox et al. [1996]; Polkovnichenko and Zhao [2013]; Baele et al. [2018]; Vandenbroucke [2015]). However, it introduces two serious tractability issues. Firstly, the non-concavity of the utility function complicates the computation of the global maximum. Secondly, the probability distortion as it is formulated in (5.17) leads to a non-additive expectation operator.

As we explain in the following section, however, the formulation we propose in this chapter includes neither of these problems; the strict convexity (concavity) of the statistical divergence (utility function) is maintained, and the probability distortion we incorporate maintains the additivity of the expectation operator.

5.4.1 The Weighted Monte Carlo Method with Probability Distortion

We now turn to the main contribution of this chapter, which is a method that combines probability distortion in style of CPT preferences with the weighted Monte Carlo method. This allows us to adapt the no-arbitrage model we want to calibrate to the thick tails implied by far-OTM options while maintaining an exact fit with the benchmark instruments we use for the calibration.

When the benchmark instruments all have the same maturity, we proceed exactly as in section 5.2 where we solve (5.5) to obtain the new weights $\boldsymbol{q}$ with the exception that we now use a prior measure $\boldsymbol{p}$ which incorporates probability distortion that re-weights the tail events of the initial model. This is in contrast to the original weighted Monte Carlo method where the prior measure is simply taken to be uniform.

We define the tail events to be large changes in the price of the underlying asset. That means we need to sort the realized paths according to their value at the time the benchmark options expire. We then enumerate them in an ascending order as $-m, -m+1, \ldots, n-1, n_1$, where path $i = 0$ is the closest to realizing no change in value for the underlying among the sampled paths,

and compute the prior measure as

$$p_i = \begin{cases} \nu^+(\frac{n-i}{N}) - \nu^+(\frac{n-i-1}{N}), & \text{if } 0 \le i \le n, \\ \nu^-(\frac{m-i}{N}) - \nu^-(\frac{m-i-1}{N}), & \text{if } -m \le i < 0, \end{cases} \tag{5.23}$$

where $\nu^+(\cdot)$ and $\nu^-(\cdot)$ are the distortion functions for the gain and loss domains, respectively.

As we mentioned in the preceding section, a portfolio choice problem with CPT preferences generally leads to a non-additive expectation operator. This is not the case in the formulation above. To understand why we avoid a non-additive expectation we recall that the Choquet integral is additive for comonotonic random variables. For a portfolio choice problem where the available instruments consist of an underlying asset and options on that asset the prospect outcomes can most generally be taken to be the value of the portfolio of the investor, which generally is not comonotonic with the underlying asset if, for example, the portfolio includes short positions on that asset. However, in our formulation the probability distortion does not depend on $\boldsymbol{\lambda}$ (or $\boldsymbol{\phi}$ in the utility maximization case), so the expectation is trivially contained within a single comonotonic class.

If the algorithm is implemented in such a way that the portfolio problem is solved for only one option maturity, the ordering is straightforward, since the value of each realized path relative to the rest is unambiguous at maturity. If more than one benchmark maturity is present, the ordering of entire paths is no longer unambiguous, since the simulated paths can cross each other between maturities. The method proposed here tackles this issue by splitting the set of benchmark instruments into single-maturity subsets, and solving the portfolio problem for each subset separately. In other words, if our benchmark instruments consist of options with M different maturities, we split the options into M groups and solve the equivalent of M single maturity entropy minimization problems. This results in each path being assigned a vector of weights. Realized values along the simulated paths at times that do not coincide with the maturities present in the set of benchmark options are then assigned weights based on linear interpolation with respect to their relative position. This is in contrast with the original weighted Monte Carlo method which assigns a single weight to entire paths.

The full specification of the method we propose is as follows. Let J_b and J_s be the two sets of indices such that $\Gamma_b = \{t_j | j \in J_b\}$ is the set of all points in time between 0 and T that coincide with the maturity of a benchmark option, and $\Gamma_s = \{t_j | j \in J_s\}$ is the set of all points in time between 0 and T that do not coincide with a benchmark maturity but for which we would like to know the state price density implied by the benchmark options. Furthermore,

given $j \in J_b$, let $S^j = \left\{x^j_{-m}, \ldots, x^j_{-1}, x^j_0, x^j_1, \ldots, x^j_n\right\}$ denote the ordered set of realized points in our sample space at time t_j, such that $x^j_i < x^j_{i'}$ for $i < i'$.

We start by computing the prior weights $\boldsymbol{p}^j$ using (5.23). Next, we compute the new weights $\boldsymbol{q}^j$ by solving either (5.5) for an exact fit or (5.6) for an approximate fit, and plugging the solution $\boldsymbol{\lambda}^j$ into (5.4). Once the weights have been computed for each maturity $t_j \in \Gamma_b$, the set of weights for each $t_{j'} \in \Gamma_s$ are calculated as follows. Assume first that we have $j, j'' \in J_b$ and $j' \in J_s$ such that $t_j = \sup\left\{\tau \in \Gamma_b | \tau < t_{j'}\right\}$ and $t_{j''} = \inf\left\{\tau \in \Gamma_s | \tau > t_{j'}\right\}$.

First, we calculate the sort index arrays I^j, $I^{j'}$, and $I^{j''}$ (i.e., the index arrays for the sorted versions of S^j, $S^{j'}$, and $S^{j''}$). Denote the weight arrays for S^j and $S^{j''}$ by $\boldsymbol{q}^j$ and $\boldsymbol{q}^{j''}$, and for the sake of visual clarity let us introduce the notation $\boldsymbol{q}^j(i) \equiv q^j_i$. If $I^{j'}(i) = i'$, we compute

$$\boldsymbol{q}^{j'}\left(i'\right) = \frac{t_{j'} - t_j}{t_{j''} - t_j} \boldsymbol{q}^j\left(I^j\left(i'\right)^{-1}\right) + \frac{t_{j''} - t_{j'}}{t_{j''} - t_j} \boldsymbol{q}^{j''}\left(I^{j''}\left(i'\right)^{-1}\right), \tag{5.24}$$

for $i = 1, \ldots, N$. Finally, if either t_j or $t_{j''}$ does not exist, i.e., if the maturity of $t_{j'}$ falls outside the range of maturities of the benchmark instruments, then we simply choose whatever benchmark maturity is closest and put the interpolation weight of that maturity to one.

To summarize, the calibration algorithm proposed above works in its general form as follows:

1. A set of N paths is generated from the initial model, and the payoff matrix in (5.1) is computed.

2. The paths are indexed according to their sort order at each of the M benchmark maturities, and the set of prior weights $\boldsymbol{p}^j$ for $j = 1, \ldots, M$ are computed.

3. With the payoffs and prior weights from 1)-2) in hand we solve separately for each t_j the problem given by (5.5) (or (5.8) for an approximate fit) using the prior weights $\boldsymbol{p}^j$ and the restriction of the payoff matrix to the set of benchmark instruments which expire at t_j. We then plug the solution $\boldsymbol{\lambda}^j$ into (5.4) to obtain $\boldsymbol{q}^j$ which become the weights attached to the points on the sample paths at time t_j.

4. For a maturity $t_{j'}$ of interest which does not coincide with the maturities present in the set of benchmark instruments we use the formula given by (5.24) to interpolate between the two weights vectors $\boldsymbol{q}^j$ and $\boldsymbol{q}^{j''}$ obtained for the two subsets of benchmark instruments whose maturities t_j and $t_{j''}$ most narrowly sandwich $t_{j'}$. In the event that

the maturity is either longer or shorter than anything that is available in the set of benchmark instruments we simply attach to it the weights vector computed for the set of benchmark instruments with the maturity closest to $t_{j'}$.

We see that if we set the probability distortion function in (5.23) equal to the identity function and drop the distinction between benchmark instruments based on their maturity then we retrieve the original weighted Monte Carlo method. For the sake of conciseness we shall hereafter refer to the procedure given above as the generalized weighted Monte Carlo (GWMC) method, and to the special case where no partition is performed on the set of benchmark instruments and where the prior measure is uniform as the original weighted Monte Carlo (OWMC) method.

5.4.2 Path Dependent Option Pricing with GWMC-Calibrated Paths

In the most basic setup of the weighted Monte Carlo method, where a single weight is assigned to each path, the probabilistic interpretation is clear. The weight q_i represents the subjective probability that path i, in the statespace that consists of the sample paths generated by the initial model, is realized. However, the GWMC method gives us several weights per path. Each weight in this case corresponds to the subjective, unconditional probability of observing a particular realization of the price of the underlying, S, at a given time t. To explain how this translates into the usual concept of Monte Carlo pricing of path dependent options, we start by giving the following theorem.

Theorem 5.1. *Let $s_i = \{s_i^1, \ldots, s_i^M\} \in \Omega$ denote the ith realization of $S = \{S^j\}_{j=1}^M$ in our sample of N paths. Furthermore, assume that for each $h, i = 1, \ldots, N$ and $j, m = 1, \ldots, M$ we have that $s_i^j = s_h^m$ if and only if $i = h$ and $j = m$, i.e., no sample path i in S contains any elements also contained by another sample path h of S. Then the probability of path i being realized under the subjective measure $\mathbb{Q}$ is given by*

$$\mathbb{Q}(\{S = s_i\}) = \frac{1}{M} \sum_{j=1}^{M} q_i^j,$$

where q_i^j is the unconditional probability under $\mathbb{Q}$ of observing s_i^j.

Proof. By the law of total probability, we have that

$$\mathbb{Q}(\{S = s_i\}) = \sum_{j=1}^{M} \mathbb{Q}\left(\left\{S = s_i | s_i^j\right\}\right) \mathbb{Q}\left(\left\{s_i^j\right\}\right),$$

where $\mathbb{Q}\left(\left\{s_i^j\right\}\right)$ is the unconditional probability that s_i^j is realized at all. We have that s_i^j completely identifies the path s_i, by our assumption of sample uniqueness. Therefore, $\mathbb{Q}\left(\left\{S = s_i | s_i^j\right\}\right) = 1$, and we can write

$$\mathbb{Q}\left(\{S = s_i\}\right) = \sum_{j=1}^{M} \mathbb{Q}\left(\left\{s_i^j\right\}\right) = \sum_{j=1}^{M} \frac{1}{M} q_i^j.$$

Note that q_i^j is the subjective probability of observing s_i^j at time t_j. However, the unconditional subjective probability of observing s_i^j at all is given by

$$\mathbb{Q}\left(\left\{s_i^j\right\}\right) = \frac{q_i^j}{M},$$

since s_i^j can only appear at time t_j, and not in any of the other $M-1$ maturities, by our assumption of unique realizations of S.

□

In other words, the subjective probability of observing a given sample path when the weights are determined by the GWMC is simply the average of the weights along that path. Note that our assumption of uniqueness among the values generated by the simulation of the initial model is a simplifying one, but we would expect it to hold for any decent random number generator.

We conclude this section by illustrating the application of the GWMC to path dependent options by the way of an example. Consider the Monte Carlo formula for the price of an arithmetic Asian option with M monitoring points. The standard Monte Carlo price for this option is given by

$$\pi_A = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{M} \sum_{t=1}^{M} s_i^j - X \right|^+,$$

with X being the strike price. Now, let q_i^j denote the weight of path i at time t_j. If we write

$$\overline{q}_i = \frac{1}{M} \sum_{j=1}^{M} q_i^j,$$

the price π_A is given by

$$\pi_A = \sum_{i=1}^{N} \overline{q}_i \left| \frac{1}{M} \sum_{j=1}^{M} s_i^j - X \right|^+.$$

If all the weights q_i^j are identical for a given i, this expression simplifies to the original weighted Monte Carlo formulation. If we further set $q_i^j = \frac{1}{N}$ for every i and every j we get the usual Monte Carlo pricing formula for the arithmetic Asian option.

5.5 Implementation Details And Numerical Results

Our numerical experiments consisted of two parts:

1. A cross-sectional run where we tested the GWMC method for single maturities only.

2. An intertemporal run where we tested the GWMC method for multiple maturities simultaneously.

In section 5.5.1 we describe the initial models we used and the pre-calibration we employed to estimate their parameters, as well as the probability distortion function we chose to implement. In sections 5.5.2 and 5.5.3 we give the numerical results from the cross-sectional and intertemporal tests, respectively.

Our numerical tests were performed on SPX options priced during the period 01/01/2013 to 31/12/2013. This data set contained a total of 765952 contracts. We calculated the risk free rate by linearly interpolating yields from US Treasury bills data available on the US Federal Reserve website. The dividend payments on the S&P 500 index were approximated by a continuous dividend yield. In addition to the options, we included the underlying asset itself as well as a risk-free asset in the set of benchmark instruments.

Throughout our empirical tests we calculated two types of error measurements. These were the mean relative price error (MRPE) and the mean average price error (MAPE). More precisely,

$$MRPE = \frac{1}{K}\sum_{k=1}^{K}\frac{|\pi_k^{\text{Model}} - \pi_k^{\text{Market}}|}{\pi_k^{\text{Market}}}, \tag{5.25}$$

and

$$MAPE = \frac{1}{K}\sum_{k=1}^{K}|\pi_k^{\text{Model}} - \pi_k^{\text{Market}}|, \tag{5.26}$$

where π_k^{Model} is the model-predicted price of benchmark instrument k and π_k^{Market} is the price of that instrument observed in the market.

The data exhibited a significant number of put-call parity violations, which can likely be attributed to the fact that we used end-of-day prices which leads to a degree of asynchronicity. For this reason all of the numerical tests are done on out-of-money puts and out-of-money calls. As the number of benchmark instruments increases, the entropy minimization/utility maximization part of the calibration procedure can become a challenge in itself, particularly when an exact solution is sought. We kept the number of benchmark instruments small for this reason, and used the least squares approach in the intertemporal tests. In addition, we performed each calibration separately for the puts and the calls to further reduce calibration failures.

For the pre-calibration (i.e., the estimation of the parameters of the geometric Brownian motion and Heston models) we used Matlab's lsqnonlin routine, and for the calibration runs we used the BFGS routine in Python's Scipy library.[27]

5.5.1 Initial Models, Pre-Calibration and Path Generation

We used two no-arbitrage models as a prior for our calibration procedure; geometric Brownian motion, and Heston's stochastic volatility model (Heston [1993]). Sampling with these models was done using the Euler discretization scheme. More specifically, for the geometric Brownian motion model we generated the paths using the discretized dynamics

$$S_{t+\Delta t} = S_t e^{(r-d-\frac{1}{2}\sigma^2)\Delta t+\sigma\sqrt{\Delta t}Z},$$

where r is the risk-free rate, d is the continuous dividend yield, σ is the volatility and Z are standard Gaussian innovations, i.e., $Z \sim N(0,1)$. For the Heston model we generated the paths using

$$\begin{aligned} S_{t+\Delta t} &= S_t e^{(r-d-\frac{1}{2}\sigma^2)\Delta t+\sqrt{v_t\Delta t}Z_S}, \\ v_{t+\Delta t} &= v_t + \kappa(\theta - v_t)\Delta t + \eta\sqrt{v_t\Delta t}Z_v, \end{aligned}$$

where S_t is the price process of the underlying asset, v_t is the variance process, r and δ are the risk free rate and dividend yield as before, κ is the rate of mean reversion, θ is the long run volatility, and η is the volatility-of-volatility. Here, the innovations $Z_v \sim N(0,1)$ and $Z_S \sim N(0,1)$ are correlated with correlation coefficient ρ.

The calibration runs for the out-of-sample performance tests were all done using $N = 40000$ simulated paths with antithetic variance reduction. For the

27. For a discussion on the computational considerations on the weighted Monte Carlo method, see Elices and Giménes [2006].

variance process in the Heston model, we used a full truncation scheme[28] for the variance process to prevent it becoming negative. Each path simulated for the Heston model contained 100 points which were distributed equally between each benchmark maturity present along the path.

The pre-calibration of these models was done using a nonlinear least squares approach. That is, the parameter values χ^* for the respective models were found by computing[29]

$$\boldsymbol{\chi}^* = \underset{\boldsymbol{\chi}}{\operatorname{argmin}} \left\{ \sum_{k=1}^{K-2} w_k \left(\pi_k^{\text{Model}} - \pi_k^{\text{Market}} \right)^2 \right\}. \tag{5.27}$$

The weights w_k, $k = 1, \ldots, K-2$ were calculated as the inverse of the bid-ask spread for the corresponding option, i.e.,

$$w_k = \frac{1}{\pi_k^{\text{Market,ask}} - \pi_k^{\text{Market,bid}}}.$$

For the Geometric Brownian model, the decision variable is $\boldsymbol{\chi} = \sigma$, whereas for the Heston model we have $\boldsymbol{\chi} = (v_0, \theta, \eta, \kappa, \rho)$. In terms of European option pricing, both the geometric Brownian motion model and the Heston model have closed form solutions, although in the latter case we make use of the characteristic function, which contains an integral which must be evaluated numerically.

A pre-calibration was done for each open market day over the period of 01/01/2013 to 31/12/2013. For each such day the set of options used in the pre-calibration procedure consisted of the 14 most traded out-of-money puts, together with the 14 most traded out-of-money calls for each maturity. For maturities where fewer than 14 puts (calls) were traded that day, we simply included all put (call) options with nonzero trading volume. The summary of the parameter estimates obtained from this pre-calibration procedure are given in Table 5.1.

5.5.2 Cross-Sectional Calibration Results

Our goal for this part of the numerical tests was to try and include as much of the strike range of the out-of-money puts and calls for each maturity as possible, to see the full effect of the over-and-underweighting of probabilities

28. Here, full truncation means the variance process is given by $\max\{V_t, 0\}$ at every sample time t.

29. Here the index only reaches $K-2$ since we performed the pre-calibration using only the option data, leaving out the underlying and risk-free assets added for the subsequent weighted Monte Carlo calibration tests.

	σ	v_0	θ	η	ρ	κ
mean	0.1313	0.0166	0.0422	0.3981	-0.9337	1.9219
std	0.0063	0.0030	0.0039	0.0313	0.0207	0.4448
max	0.1435	0.0232	0.0477	0.4655	-0.8979	2.8056
min	0.1237	0.0127	0.0341	0.3514	-0.9715	1.3233

Table 5.1: Pre-Calibration summary for initial models.

of extreme events on the pricing measure. However, at the extreme ends of option moneyness the data becomes noticeably less reliable, with instances of duplicate prices for options with different strikes, bid prices of zero, and so on. For these reasons, we removed the following:

- Any option with bid price zero.
- Any option with price lower than 0.5
- Any set of options with different strikes but the same price, same option type, and same maturity quoted on the same day.
- Any option which's open interest count fell below 2000 for short and medium maturities, and 1000 for long maturities.
- Any set of options of the same maturity quoted on a given trading day with fewer than 10 options satisfying the above criteria.

Here, short maturities are defined to be between 1 and 90 days, medium maturities between 91 and 250 days, and long maturities anything longer than 250 days. We used open interest as our measure of liquidity, instead of trading volume, since trading volume gave a much thinner support for options at the far ends of the moneyness spectrum, as well as options with long maturities.

After the aforementioned contracts had been removed, we were left with a set of 24263 contracts, consisting solely of out-of-money puts and calls. For this set of contracts we performed an out-of-sample test for 3 different calibration methods for 2 different initial models:

1. The unweighted geometric Brownian motion model.
2. The unweighted Heston model.
3. The geometric Brownian motion model calibrated with the OWMC.

4. The Heston model calibrated with the OWMC.
5. The geometric Brownian motion model calibrated with the GWMC.
6. The Heston model calibrated with the GWMC.

For each model-method pair we calculated the out-of-sample performance for each maturity and option type (i.e puts, calls) separately, as well as their aggregate performance over these categories, and these results are given in Tables 5.4 and 5.5, while Table 5.3 gives the aggregate result for each model-method pair. The first entry in each number pair in these tables is the mean relative pricing error, and the second entry is the mean absolute pricing error, as explained in section 5.5.1. For each option type (i.e., put or call), we used as benchmark instruments the 5 options with strikes closest to the forward price of the underlying, so a total of 10 benchmark options per maturity. The remainder of the options in the data set, obtained from the data cleaning procedure described above, were used as out-of-sample instruments.

For the calibration of a given trading day, we used the distortion parameter values, $\boldsymbol{\xi} = \{\gamma^+, \delta^+, \gamma^-, \delta^-\}$, which gave the best out-of-sample performance during the calibration of the previous trading day.[30] More specifically, before calibrating the model for a given trading day, we solve the following for the trading day immediately before it:

$$\boldsymbol{\xi}^* = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \left\{ \frac{1}{K} \sum_{k=1}^{K} \frac{|\pi_k^{\text{Model}}(\boldsymbol{\xi}) - \pi_k^{\text{Market}}|}{\pi_k^{\text{Market}}} \right\}. \tag{5.28}$$

This optimization problem was solved using BFGS. While each function evaluation in this minimax optimization problem is expensive (we are solving the portfolio choice problem with each function call) the computational times are drastically reduced by reusing the previous optimal parameter estimate as a starting point, since the optimal values turn out to change very little between trading days in our data. In addition, since there is no interaction between the gain domain parameters and the loss domain parameters, the objective function is separable into two parts, each with only two decision variables (i.e., the elevation and curvature parameters), which further reduces the computational load.

30. As can be seen from comparing the values in Table 5.2, we expect these values to be different for different initial models.

Model	Horizon	γ^+	δ^+	γ^-	δ^-
GBM	Short	1.137	0.921	0.431	0.577
		(0.124)	(0.132)	(0.129)	(0.178)
	Medium	1.051	0.975	0.316	0.611
		(0.138)	(0.166)	(0.111)	(0.142)
	Long	1.100	1.022	0.332	0.580
		(0.083)	(0.090)	(0.105)	(0.107)
Heston	Short	0.614	0.588	0.630	0.547
		(0.149)	(0.168)	(0.153)	(0.177)
	Medium	0.555	0.623	0.791	0.640
		(0.157)	(0.182)	(0.139)	(0.210)
	Long	0.451	0.601	0.788	0.754
		(0.112)	(0.124)	(0.116)	(0.118)

Table 5.2: The averages of the probability distortion parameter values for different maturity clusters and models, along with their standard deviations in brackets.

Method	GBM	Heston
Unweighted	0.736, 9.727	0.313, 3.316
OWMC	0.378, 2.439	0.220, 0.796
GWMC	0.174, 0.643	0.066, 0.286

Table 5.3: Aggregate results for the out-of-sample cross-sectional performance for the 2 different initial models along with the 2 different weighted calibration methods, as well as the unweighted versions for comparison. The first and second number in each column are the average MRPE and MAPE, respectively, for the entire data set.

Method	Strike	Short	Medium	Long	All
MC	K<F	0.624, 3.216	0.818, 9.492	0.853, 24.52	0.748, 11.286
	K>F	0.855, 3.515	0.53, 3.429	0.345, 3.63	0.69, 3.513
	All	0.684, 3.294	0.768, 8.442	0.789, 21.901	0.736, 9.727
OWMC	K<F	0.439, 1.071	0.506, 2.444	0.444, 5.625	0.457, 3.042
	K>F	0.121, 0.275	0.089, 0.295	0.076, 0.438	0.101, 0.331
	All	0.355, 0.86	0.435, 2.078	0.369, 4.567	0.378, 2.439
GWMC	K<F	0.186, 0.331	0.224, 0.567	0.208, 1.343	0.203, 0.755
	K>F	0.072, 0.155	0.079, 0.264	0.068, 0.391	0.072, 0.249
	All	0.155, 0.284	0.198, 0.514	0.180, 1.154	0.174, 0.643

Table 5.4: Cross-sectional results for the unweighted Monte Carlo method (MC), the OWMC method and the GWMC method using geometric Brownian motion as an initial model, for OTM puts (K<F) and OTM calls (K>F). The first and second number in each column are the MRPE and MAPE, respectively.

Method	Strike	Short	Medium	Long	All
MC	K<F	0.234, 0.798	0.362, 2.861	0.391, 8.284	0.317, 3.592
	K>F	0.292, 1.569	0.314, 2.356	0.297, 4.283	0.298, 2.215
	All	0.249, 1.001	0.353, 2.773	0.379, 7.782	0.313, 3.316
OWMC	K<F	0.271, 0.544	0.222, 0.78	0.152, 1.138	0.216, 0.815
	K>F	0.216, 0.397	0.27, 0.813	0.239, 1.21	0.232, 0.73
	All	0.256, 0.505	0.23, 0.787	0.17, 1.153	0.22, 0.796
GWMC	K<F	0.084, 0.206	0.049, 0.195	0.067, 0.593	0.067, 0.322
	K>F	0.075, 0.147	0.044, 0.120	0.042, 0.161	0.060, 0.142
	All	0.081, 0.190	0.048, 0.182	0.063, 0.536	0.066, 0.286

Table 5.5: Cross-sectional results for the unweighted Monte Carlo method (MC), the OWMC method and the GWMC method using the Heston model as an initial model, for OTM puts (K<F) and OTM calls (K>F). The first and second number in each column are the MRPE and MAPE, respectively.

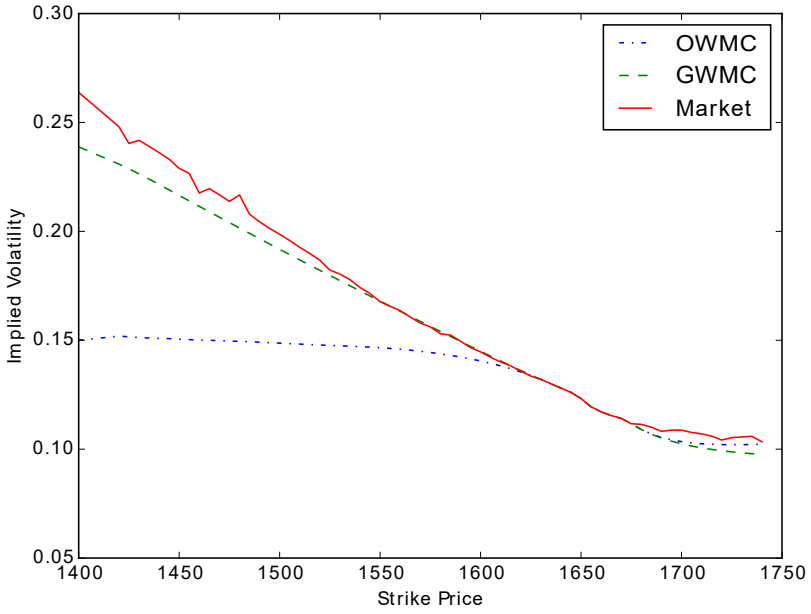

Figure 5.3: The volatility for SPX options with maturity 30 days implied by a) market prices (solid line), b) the OWMC calibrated GBM model (dashed line), and c) the GWMC calibrated GBM model (dot-dashed line).

Figures 5.3 and 5.4 give an idea of how the GWMC improves upon the original weighted Monte Carlo method in terms of the empirical fit. They show a cross-section of the implied volatility calculated from market prices of options traded on the 21st of May 2013 with maturity of 30 days, plotted together with the OWMC and GWMC implied volatilities for those same options. The benchmark options for the OWMC and GWMC consisted of 5 close to the money puts and 5 close to the money calls. The figures demonstrate a typical difference between the OWMC and the GWMC as pricing models; the former tends to underprice far-out-of-money options to a much greater degree than the latter, with the exception of call options for geometric Brownian motion as the initial model.

When comparing the performance of a given model across the three maturity categories it should be kept in mind that the presence of options of extreme moneyness tended to be significantly more prevalent in the medium range of our data than in the other two maturity groups, in addition to the fact that the liquidity requirements were not the same across the maturity

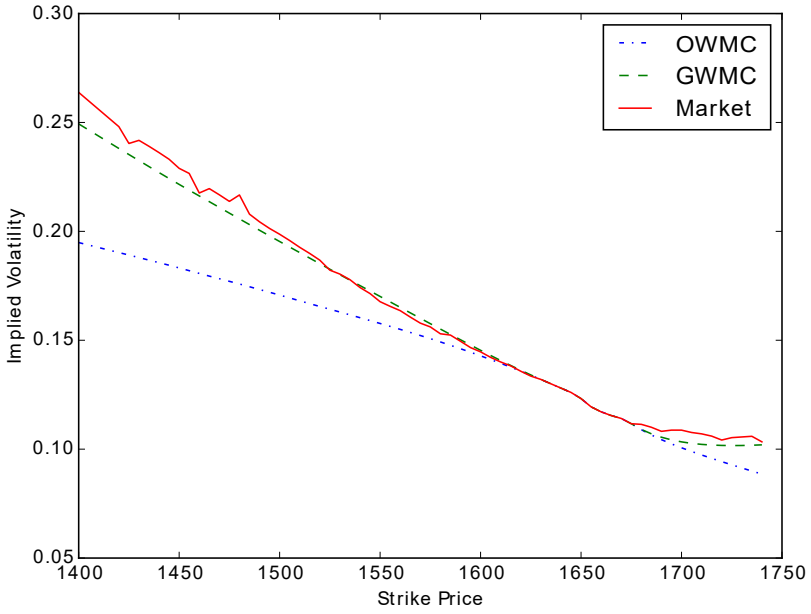

Figure 5.4: The volatility for SPX options with maturity 30 days implied by a) market prices (solid line), b) the OWMC calibrated Heston model (dashed line), and c) the GWMC calibrated Heston model (star-dashed line).

groups. With that said, the Heston model outperforms the GBM model overall as expected in the unweighted case. Interestingly, though, the OWMC for GBM appears to achieve very good performance for the call options, which highlights the fact that a poor empirical fit of the initial model does not necessitate a poor fit once the paths have been reweighted. Overall, however, the GWMC method produces significant improvements, both for the geometric Brownian motion and the Heston model compared to the OWMC. It should also be noted here that one reason why the weighted GBM models underperform, particularly for the put options, is that the Monte Carlo simulation often did not produce any paths that reached the extreme levels necessary for the most far-out options to become exercised, while the Heston model did. And the distortion function does not take effect for zero-measure events.

5.5.3 Intertemporal Calibration Results

The numerical tests in this section consisted of calibrating the initial models of section 5.5.1 to benchmark options spanning more than one maturity. For each benchmark option maturity present during a given trading day, we chose the 10 out-of-money put options and the 10 out-of-money call options with the highest trading volume, and out of these we used the 5 out-of-money put options and the 5 out-of-money call options that had strikes closest to the forward value as benchmark options. The remainder of the options served as out-of-sample options. We used the same probability distortion values as are given in Table 5.2 in section 5.5.2.

For the purpose of demonstration, let U_d, with $d = 1,\ldots,D$ denote the set of all options for trading day d, where D is the final day (i.e., 31st of December 2013), in our sample with nonzero trading volume. Furthermore, for each U_d let $M_{d,j} \subseteq U_d$ denote the set of the 10 puts and 10 calls with the highest trading volume with maturity m_j, where the maturities are indexed here by $j = 1,\ldots,J$ in such a way that if $j' < j''$ then $m_{j'} < m_{j''}$. Also, let $K_{d,j} \subseteq M_{d,j}$ be the set of 5 puts and 5 calls in $M_{d,j}$ that are closest to being at-the-money. Lastly, define the Cartesian product $H_d := \otimes_{j=1}^{J} M_{d,j}$, and let $\mathrm{Dim}(H_d) := J$, be the number of different maturities present in H_d. The intertemporal test can then be described in the following way. For a given trading day d:

1. if $\mathrm{Dim}(H_d) \leq 3$ that trading day is dropped from the sample,

2. if $\mathrm{Dim}(H_d) = 4$ the in-sample instruments consist of $K_{d,1}$ and $K_{d,3}$, and the out-of-sample instruments are $M_{d,1}\backslash K_{d,1}$, $M_{d,2}$, $M_{d,3}\backslash K_{d,3}$ and $M_{d,4}$,

3. if $\mathrm{Dim}(H_d) = 5$ the in-sample instruments consist of $K_{d,2}$ and $K_{d,4}$, and the out-of sample instruments are $M_{d,1}$, $M_{d,2}\backslash K_{d,2}$, $M_{d,3}$, $M_{d,4}\backslash K_{d,4}$ and $M_{d,5}$,

4. if $\mathrm{Dim}(H_d) \geq 6$ the in-sample instruments consist of $K_{d,2+\iota}$ and $K_{d,5+\iota}$, and the out-of-sample instruments are $M_{d,1+\iota}$, $M_{d,3+\iota}\backslash K_{d,3+\iota}$, $M_{d,3+\iota}$, $M_{d,4+\iota}$, $M_{d,5+\iota}\backslash K_{d,5+\iota}$, and $M_{d,6+\iota}$, with $\iota = 0,\ldots,\mathrm{Dim}(H_d) - 6$.

To elaborate on part 4, for H_d with $\mathrm{Dim}(H_d) = J > 6$ we begin by setting $\iota = 0$ and calibrate. Once the calibration is done and we have calculated the out-of-sample prices, if $\mathrm{Dim}(H_d) > 6$, we set $\iota = 1$ which shifts the maturity index by one to the right to perform the calibration and out-of-sample calculation again, and so on, until $M_{d,6+\iota} = M_{d,J}$. This test scheme resulted in the calculation of 135792 option prices.[31]

Method	Strike	MRPE	MAPE
OWMC	K<F	0.313	1.859
	K>F	0.267	1.163
	All	0.303	1.716
GWMC	K<F	0.151	0.899
	K>F	0.089	0.490
	All	0.136	0.816

Table 5.6: Intertemporal run results for the OWMC method and the GWMC method using geometric Brownian motion as an initial model for OTM puts (K<F) and OTM calls (K>F).

Method	Strike	MRPE	MAPE
OWMC	K<F	0.164	0.692
	K>F	0.202	0.696
	All	0.172	0.693
GWMC	K<F	0.074	0.401
	K>F	0.103	0.357
	All	0.080	0.392

Table 5.7: Intertemporal run results for the OWMC method and the GWMC method using the Heston model as a initial model for OTM puts (K<F) and OTM calls (K>F).

The design of the intertemporal test was made with the aim of avoiding biases in the selection of the benchmark options by including an approximately even mix of "outside" and "inbetween" options, and by rolling over every maturity as in step 4 for each trading day. As the results in Tables 5.6 and 5.7 show, the Heston model gives better results overall than the geometric Brownian motion, and the GWMC likewise improves upon the OWMC in all categories. The only break from this pattern of improvement is the MRPE for the OTM call category for the GWMC, which is smaller for the GBM model than the Heston model which we can trace back to the excess volatility on the up side of the underlying for the geometric Brownian motion, which counters the exponential decay of the tails that we get from using the Kullback-Leibler divergence more aggressively.

31. Most trading days fell under the fourth case, meaning most options were priced several times using different benchmarks, and the number reflects every such calculation.

5.6 Conclusion

We have presented a generalization of the weighted Monte Carlo calibration method first proposed by Avellaneda et al. [2001] which we show yields a significant improvement in empirical fit through extensive numerical testing using S&P 500 options data. The method is centered around relative entropy minimization like the original method, but instead of assuming a uniform prior measure for the sample paths we compute a measure which distorts the probability of tail events. This allows us to overcome the consistent implied volatility gap between the market data and what the original WMC method produces. Furthermore, the proposed method partitions the set of benchmark instruments by maturity and assigns multiple weights per path which increases the flexibility of the method, as opposed to the original method which assigns only one weight per path, and can be considered a special case of what we have presented here.

Furthermore, we have provided a detailed discussion on the utility maximization problem which is known to be dual to the relative entropy minimization formulation. The utility maximization formulation presented in this chapter is technically speaking more general than the relative entropy minimization formulation, and fits with any type of preferences which are of expected utility type which opens up the possibility of refinements of the weighted Monte Carlo method through a purely utility-based approach.

Bibliography

Abramowitz, M., & Stegun, I. A. (1964). Error Function and Fresnel Integrals. *Handbook of Mathematical Functions* (pp. 295-300). National Bureau of Standards, Washington, DC.

Abramowitz, M. Stegun, I (1964). Confluent Hypergeometric Functions, *Handbook of Mathematical Functions* (pp. 503-537). National Bureau of Standards, Washington, DC.

Aichinger, M. & Binder, A. (2013). Characteristic Function Methods for Option Pricing. *A Workout in Computational Finance* (pp. 193-209). John Wiley & Sons.

Aït-Sahalia, Y. & Kimmel, R. (2007). Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics*, 83:413–452

Albanese, C., Jackson, K. & Wiberg, P. (2004). A new Fourier transform algorithm for value-at-risk. *Quantitative Finance*, 4(3):328-338.

Albrecher, H., Mayer, P., Schoutens, W. & Tistaert, J. (2007). The Little Heston Trap. *Wilmott Magazine*, (Jan):83-92.

Arismendi, J. C., Back, J., Prokopczuk, M., Paschke, R. & Rudolf, M. (2016). Seasonal stochastic volatility: Implications for the pricing of commodity options. *Journal of Banking & Finance*, 66:53-65.

Attari, M. (2004). *Option pricing using Fourier transforms: A numerically efficient simplification* (Working Paper). Available at SSRN: `https://ssrn.com/abstract=520042` or `http://dx.doi.org/10.2139/ssrn.520042`

Avellaneda, M., Buff, R., Friedman, C., Grandechamp, N., Kruk, L. & Newman, J. (2001). Weighted Monte Carlo: a new technique for calibrating asset-pricing models. *International Journal of Theoretical and Applied Finance*, 4(01):91-119.

Bank For International Settlements, *Statistical Release: OTC Derivative Statistics at end-December 2018* (pp. 1-9). Available at `https://www.bis.org/publ/otc_hy1905.htm`

Back, J., Prokopczuk, M. & Rudolf, M. (2013). Seasonality and the valuation of commodity options. *Journal of Banking & Finance*, 37(2):273-290.

Back, K. (2010). *Asset pricing and portfolio choice theory*. Oxford University Press. New York, NY.

Baele, L., Driessen, J., Ebert, S., Londoño, J. M., & Spalt, O. G. (2018). Cumulative prospect theory, option returns, and the variance premium. The Review of Financial Studies, 32(9), 3667-3723.

Bakshi, G., Ju, N. & Ou-Yang, H. (2006). Estimation of continuous-time models with an application to equity volatility dynamics. *Journal of Financial Economics*, 82(1):227-249.

Bams, D., Lehnert, T. & Wolff, C. (2009). Loss Functions in Option Valuation: A Framework for Model Selection. *Management Science*, 55(5):853-862.

Barberis, N. (2013). The psychology of tail events: Progress and challenges. *The American Economic Review*, 103(3):611-616

Barone-Adesi, G. & Whaley, R. E. (1987). Efficient analytic approximation of American option values. *The Journal of Finance*, 42(2):301-320.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *The Review of Financial Studies*, 9(1):69-107.

Bates, D. S. (2003). Empirical option pricing: A retrospection. *Journal of Econometrics*, 116(1-2):387-404.

Biagini, F. (2010). Second fundamental theorem of asset pricing. In R. Cont, (Eds). *Encyclopedia of Quantitative Finance* (pp. 1623–1628). Wiley.

Black, F. (1976). The pricing of commodity contracts. Journal of financial economics, 3(1-2), 167-179.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637-654.

Brandt, M. W. & Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 63(2):161-210.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA Journal of Applied Mathematics*, 6(1):76-90.

Burkovska, O., Gaß, M., Glau, K., Mahlstedt, M., Schoutens, W., & Wohlmuth, B. (2018). Calibration to American options: numerical investigation of the de-Americanization method. Quantitative finance, 18(7):1091-1113.

Camerer, C. F. (2004). Prospect theory in the wild: Evidence from the field. *Advances in Behavioral Economics* (pp. 148-161). Princeton University Press.

Carmona, R. (2009). *Indifference pricing: theory and applications*. Princeton University Press.

Carr, P., & Madan, D. (1999). Option valuation using the fast Fourier transform. *Journal of computational finance*, 2(4):61-73.

Carr, P., & Sun, J. (2007). A new approach for option pricing under stochastic volatility. *Review of Derivatives Research*, 10(2):87-150.

Cezaro, A. D., Scherzer, O. & Zubelli, J. P. (2010). A convex-regularization framework for local volatility calibration in derivative markets: The connection with convex risk measures and exponential families. *Sixth World Congress of the Bachelier Finance Society*.

Chen, M. Y. & Chen, B. T. (2014). Online fuzzy time series analysis based on entropy discretization and a Fast Fourier Transform. *Applied Soft Computing*, 14:156-166.

Chernov, M., & Ghysels, E. (2002). A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. *Journal of Financial Economics*, 56:407–458.

Chesney, M., Jeanblanc, M. & Yor, M. (2009). *Mathematical methods for financial markets*. Springer Science & Business Media.

Choi, J. W. & Longstaff, F. A. (1985). Pricing options on agricultural futures: An application of the constant elasticity of variance option pricing model. *Journal of Futures Markets*, 5(2):247-258.

Chourdakis, K. (2005). Option pricing using the fractional FFT. *Journal of Computational Finance*, 8(2):1-18.

Christoffersen, P. & Jacobs, K. (2004). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291-318.

Cochrane, J. (2000). *Asset Pricing*. Princeton University Press

Cont, R. & Tankov, P. (2004). Non-parametric Calibration of Jump-diffusion Option Pricing Models. *Journal of Computational Finance*, 7(3):1–49.

Cont, R. (2010). Model calibration. *Encyclopedia of Quantitative Finance* (pp. 1210–1218). John Wiley & Sons.

Crawford, G., & Sen, B. (1996). *Derivatives for decision makers: Strategic management issues (Vol 1)*. John Wiley & Sons.

Crépey, S. (2010). Tikhonov Regularization. *Encyclopedia of Quantitative Finance* (pp. 1807–1812). John Wiley & Sons.

Cui, Y., del Baño Rollin, S., & Germano, G. (2017). Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research*, 263(2):625-638.

del Baño Rollin, S., Ferreiro-Castilla, A., & Utzet, F. (2010). On the density of log-spot in the Heston volatility model. *Stochastic Processes and their Applications*, 120(10):2037-2063.

Downarowicz, A. (2010). The first fundamental theorem of asset pricing. *Revista de Economía Financiera*, 21:23-35.

Drimus, G. G. (2012). Options on realized variance by transform methods: a non-affine stochastic volatility model. *Quantitative Finance*, 12(11):1679-1694.

Duffy, D. J. (2013). *Finite Difference methods in financial engineering: a Partial Differential Equation approach*. John Wiley & Sons.

Egger, H. & Engl, H. W. (2005). Tikhonov regularization applied to the inverse problem of option pricing: convergence analysis and rates. *Inverse Problems*, 21(3):1027-1045.

Elices, A. & Giménes, E. (2006). Weighted Monte Carlo. *Risk Magazine*, 19:78-83.

Eraker, B., Johannes, M. & Polson, N. (2003). The impact of jumps in volatility and returns. *The Journal of Finance*, 58(3):1269-1300.

Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance*, 59:1367–1404.

Fama, E. F. & French, K. R. (1987). Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *Journal of Business*, (Jan):55-73.

Fang, F. & Oosterlee, C. W. (2008). A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 31(2):826-848.

Fox, C. R., Rogers, B. A. & Tversky, A. (1996). Option Traders Exhibit Subadditive Decision Weights. *Journal of Risk and Uncertainty*, 13:5-17.

Friedman, C. A., Cao, W., Huang, Y. & Zhang, Y. (2013). *Engineering More Effective Weighted Monte Carlo Option Pricing Models* (Working Paper). Available at SSRN: `https://ssrn.com/abstract=2193807` or `http://dx.doi.org/10.2139/ssrn.2193807`

Friedman, C. A., Cao, W. & Huang, Y. (2014). *Some Economically Meaningful Option Model Calibration Performance Measures* (Working Paper). Available at SSRN: `https://ssrn.com/abstract=2193803` or `http://dx.doi.org/10.2139/ssrn.2193803`

Fusai, G., Germano, G. & Marazzina, D. (2016). Spitzer identity, Wiener-Hopf factorization and pricing of discretely monitored exotic options. *European Journal of Operational Research*, 251(1):124-134.

Gaß, M., Glau, K., & Mair, M. (2017). Magic points in finance: Empirical integration for parametric option pricing. SIAM Journal on Financial Mathematics, 8(1):766-803.

Gerlich, F., Giese, A. M., Maruhn, J. H. & Sachs, E. W. (2012). Parameter identification in financial market models with a feasible point SQP algorithm. *Computational Optimization and Applications*, 51(3):1137-1161.

Gilli M. & Schumann E. (2011). Calibrating Option Pricing Models with Heuristics. In: Brabazon A., O'Neill M. & Maringer D. (Eds), *Natural Computing in Computational Finance* (pp. 9-37). Springer, Berlin, Heidelberg.

Glasserman, P. (2013). *Monte Carlo methods in financial engineering*. Springer Science & Business Media. New York, NY.

Grasselli, M. (2017). The 4/2 stochastic volatility model: a unified approach for the Heston and the 3/2 model. *Mathematical Finance*, 27(4):1013-1034.

Griewank, A., & Walther, A. (2008). *Evaluating derivatives: principles and techniques of algorithmic differentiation*. Society for Industrial and Applied Mathematics.

Gudmundsson, H., & Vyncke, D. (2019). On the Calibration of the 3/2 Model. *European Journal of Operational Research*, 276(3):1178-1192.

Gudmundsson, H., & Vyncke, D. (2019). *Non-Affine Stochastic Volatility with Seasonal Trends* (Working paper).

Gudmundsson, H., & Vyncke, D. (2019). *A Generalized Weighted Monte Carlo Calibration Method* (Working paper).

Guillaume, F. & Schoutens, W. (2010). Use a reduced Heston or reduce the use of Heston?. *Wilmott Journal*, 2(4):171-192.

Gupta, A., & Reisinger, C. (2014). Robust calibration of financial models using Bayesian estimators. *J. Comput. Finance*, 17:3-36.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2):327-343.

Hirsa, A. (2016). Model Calibration. In M.A.H. Dempster, D.B. Madan, & R. Cont (Eds.), *Computational methods in finance* (pp. 259-341). CRC Press. London, UK.

Jäckel, P. & Kahl, C. (2005). Not-so-complex logarithms in the Heston model. *Wilmott magazine*, 19(9):94-103.

Johannes, M. & Polson, N. (2009). MCMC methods for continuous-time financial econometrics. *Handbook of Financial Econometrics*, 2(1):1-72.

Jones, C. S. (2003). The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics*, 116(1-2):181-224.

Kahneman, D. & Tversky, A. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5:297-323.

Kilin, F. (2011). Accelerating the calibration of stochastic volatility models. *The Journal of Derivatives*, 18(3):7-16.

Koch, K. R. (2007). *Introduction to Bayesian statistics*. Springer Science & Business Media. New York, NY.

Koekebakker, S. & Lien, G. (2004). Volatility and price jumps in agricultural futures prices—evidence from wheat options. *American Journal of Agricultural Economics*, 86(4):1018-1031.

Kullback, S. & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79-86.

Li, H., Wells, M. T. & Yu, C. L. (2006). A Bayesian analysis of return dynamics with Lévy jumps. *The Review of Financial Studies*, 21(5):2345-2378.

Ludvigson, S. C. (2011). *Advances in consumption-based asset pricing: Empirical Tests*(No. w16810). National Bureau of Economic Research. Cambridge, MA.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431-441.

Mikhailov, S. & Nögel, U. (2004). Heston's stochastic volatility model: Implementation, calibration and some extensions. *The Best of Wilmott* (pp. 401-411). John Wiley and Sons.

Milstein, G. N. (1975). Approximate integration of stochastic differential equations. *Theory of Probability & Its Applications*, 19(3):557-562.

Morozov, V. A. (1966). On the solution of functional equations by the method of regularization. *Doklady Mathematics*, 7(1):414-417.

Nau, R. F., Jose, V. R. R. & Winkler, R. L. (2009). Duality between maximization of expected utility and minimization of relative entropy when probabilities are imprecise. *ISIPTA*, 9:337-346.

Nocedal, J. & Wright, S. (2006). *Numerical optimization*. Springer Science & Business Media. New York, NY.

Pan, J. (2002). The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3-50.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55-71.

Polkovnichenko, V. & Zhao, F. (2013). Probability Weighting Functions Implied in Option Prices. *Journal of Financial Economics*, 107:580-609.

Poteshman, A. M. (1998). *Estimating a general stochastic variance model from option prices* (Working Paper). University of Chicago.

Prelec, D. (1998). The Probability Weighting Function. *Econometrica*, 66:497-527.

Storn, R. & Price, K. (1997). Differential evolution: a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341-359.

Richter, M. & Sørensen, C. (2002). *Stochastic volatility and seasonality in commodity futures and options: The case of soybeans* (Working paper). Copenhagen Business School.

Samperi, D. (1997). *Inverse Problems, Model Selection and Entropy in Derivative Security Pricing* (PhD thesis). New York University.

Schmelzle, M. (2010). *Option pricing formulae using Fourier transform: Theory and application* (Working Paper). Available at `http://pfadintegral.com`.

Schneider, L. & Tavin, B. (2015). *Seasonal Stochastic Volatility and Correlation together with the Samuelson Effect in Commodity Futures Markets* (Preprint arXiv:1506.05911). Available at `https://arxiv.org/abs/1506.05911`.

Schoutens, W., Simons, E. & Tistaert, J. (2003). A perfect calibration! Now what?. *The best of Wilmott*, 2:281-305.

Sepp, A. (2008). Pricing Options on Realized Variance in Heston Model with Jumps in Returns and Volatility. *Journal of Computational Finance*, 11(4):33-70.

Skiadas, C. (2009). *Asset pricing theory*. Princeton University Press. Princeton, NJ.

Sørensen, C. (2002). Modeling seasonality in agricultural commodity futures. *Journal of Futures Markets*, 22(5):393-426.

Tikhonov, A. N., Arsenin, V. I. & John, F. (1979). Solutions of ill-posed problems. *SIAM Rev.*, 21(2):266-267.

Trolle, A. B. & Schwartz, E. S. (2009). Unspanned stochastic volatility and the pricing of commodity derivatives. *The Review of Financial Studies*, 22(11):4423-4461.

Vandenbroucke, J. (2015). A cumulative prospect view on portfolios that hold structured products. *Journal of Behavioral Finance*, 16(4):297-310.

Wong, H. Y. & Lo, Y. W. (2009). Option pricing with mean reversion and stochastic volatility. *European Journal of Operational Research*, 197(1):179-187.

Yang, S. & Lee, J. (2012). Multi-basin particle swarm intelligence method for optimal calibration of parametric Lévy models. *Expert Systems with Applications*, 39(1):482-493.