Delft University of Technology
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft Institute of Applied Mathematics

Lévy processes and Option Pricing

A thesis submitted to the
Delft Institute of Applied Mathematics
in partial fulfillment of the requirements

for the degree

MASTER OF SCIENCE
in
APPLIED MATHEMATICS

by

STEVEN P. TEN HAVE

Delft, the Netherlands
June 2012

MSc THESIS APPLIED MATHEMATICS

"Lévy processes and Option Pricing"

Steven P. ten Have

Delft University of Technology

Responsible professor

Prof. dr. ir. C.W. Oosterlee

Daily supervisors

Dr. S.F. Kerstan (Optiver)

Dr. I.O. Petursson (Optiver)

Other thesis committee members

Dr. N. Budko

Dr. J.A.M. van der Weide

June 2012

Delft, the Netherlands

Preface and acknowledgements

This thesis was created over the past 9 months at Optiver, a market making and proprietary trading firm in Amsterdam. The theory of Lévy processes was introduced to me by Kees Oosterlee's course on Computational Finance, which he teaches in Delft. The initial goal of this project was simple: to find a suitable Lévy process that fits market prices of options. I was placed on the trading floor of Optiver in a trading group, having regular meetings on my progress with Kees and my tutors at Optiver, Ingi Petursson and Sven Kerstan.

Initially the goal was somewhat open-ended: we were of course not guaranteed to find any positive results. We stated some backup plan for if this would fail. Fortunately however this all went quite well: within a reasonable amount of time I had adjusted the CGMY model somewhat to obtain desirable results. The project went into an area we had not so much anticipated: the actual implementation of the pricing tool began dominating the theoretical work on the stochastic processes, with as a side result me learning a new programming language (C++) to implement the pricing tools.

Personally I have experienced the setup of this project as very beneficial for myself. I have had access to traders and researchers at Optiver to ask them almost anything about mathematical finance, which was really helpful to gain knowledge about practical issues in finance.

I would very much like to thank Kees Oosterlee for the effort he put in my guidance. First of all for his expertise in pricing derivatives using the COS-method, and second for being a kind and passionate person that always has an idea for a new direction in research. Hans van der Weide deserves a word of thank for guiding me through the fine details of the stochastic processes involved. Furthermore I wish to thank Sven Kerstan and Ingi Petursson for so many fruitful discussions on my thesis topic, and finance in general. It was a truly creative period, with many (sometimes clashing) ideas and open minded discussions. It would have been hard for me to find a place where I could have learnt so much in so little time, elsewhere.

At last I wish to thank my parents, for having been there for me, supporting me in whatever I have done and will do. I think my ability to think freely and to be creative, is a direct consequence of that. My girlfriend Laura also deserves a special mention, as she has the special gift of putting a smile on my face, always. I am looking forward to life with you, after this thesis.

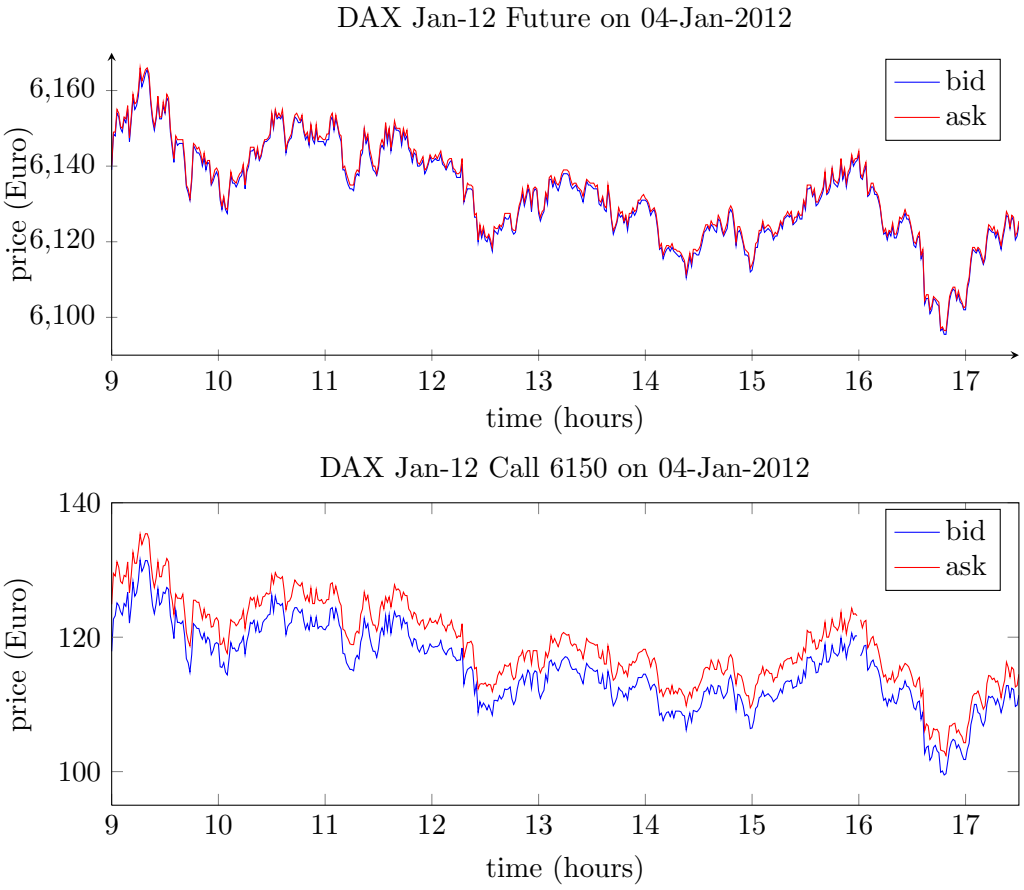
Steven ten Have

Contents

Chapter 1

Introduction: from Black-Scholes to Lévy processes

The main objective of this thesis is to find a stochastic process in the family of Lévy processes that fits prices of German DAX index options. Specifically, we are interested in options of shorter maturities (shorter than three months), as these are not modelled well with more conventional models such as Heston or SABR. With a fit, we mean to reproduce market prices within the observed bid and ask prices within a consistent model. We will analyse some known Lévy models, such as the Black-Scholes model, Variance Gamma, NIG, Meixner and CGMY. We will extend the CGMY model in such a way that it fits the observed prices better, and will call this the Augmented CGMY model.



In the above picture we see bid and ask prices for the front month DAX future expiring in January 2012 and the call option expiring in the same month with strike price 6150. We will explain the meaning of these products in a moment. Notice that the movement of the prices is related. Interestingly the call option moves much faster than the future: while the future loses about 1% of its value, the call option loses around 20%. We now turn to the pricing of options.

1.1 The Black-Scholes model

The Black-Scholes model (Black and Scholes [1973]) defines the dynamics of a stock price as a *geometric Brownian motion* (GBM):

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \tag{1.1}$$

This means the stock price S_t increases with a drift μ over time, and a 'variance' term is introduced by the factor σ, the *volatility* of the process. The process W_t is Brownian motion. Equation (1.1) is a stochastic differential equation (for an introduction to stochastic calculus, see Steele [2001]), that has an analytical solution

$$S_t = \exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right). \tag{1.2}$$

We introduce the concept of a European call option:

Definition 1. *A European call option is a contract that gives the holder the right (but not the obligation) to buy an asset at a certain strike price K at some time of maturity T.*

The payoff of a call option at time T is thus $(S_T - K)^+$. A put option is constructed exactly the same, except the holder has the right to *sell* the asset for the strike price K. The payoff for the European put is thus $(K - S_T)^+$.

The key question is of course: what should be the *price* of an option at any given time t? To answer this question, the concept of *risk-neutral pricing* has been developed that rests on martingale theory (see Shreve [2005]). We will explain it in the next section.

For the Black-Scholes model, the price of a European call option is known analytically as

$$C(S,t) \quad = \quad N(d_1)S - N(d_2)Ke^{-r(T-t)} \tag{1.3}$$

where

$$\begin{aligned} d_1 &= \frac{\log(S/K) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}, \\ d_2 &= d_1 - \sigma\sqrt{T-t}. \end{aligned}$$

This formula (introduced in Black and Scholes [1973] and Merton [1973]) revolutionized the theory of (mathematical) finance. Robert C. Merton and Myron Scholes received the Nobel Prize in Economics for it in 1997 (Fischer Black had died in 1995). The model however exhibits some severe shortcomings, which we shall discuss throughout this thesis.

1.2 The Put-Call Parity

The Put-Call Parity is a general relationship between the prices of European put and call options. Consider two portfolio's set up at some time t. The first portfolio buys a European call option with strike K and expiration T for a price $C(t)$, and writes a European put option with strike K and expiration T at a price $P(t)$. A second portfolio is set up, buying one share at price $S(t)$ and borrow K bonds at price $KB(t)$.

The payoff of both portfolios at time T is $S(T) - K$. Therefore, at time t one should be indifferent between setting up the first or the second portfolio. This implies that the value of the first portfolio should always be equal to the value of the second portfolio:

$$C(t) - P(t) = S(t) - KB(t). \tag{1.4}$$

If we furthermore assume a constant interest rate r, we find that the price of a bond that pays one at time T must be $B(t) = e^{-r(T-t)}$. The Put-Call Parity then becomes

$$C(t) - P(t) = S(t) - Ke^{-r(T-t)}. \tag{1.5}$$

It is important to realize that this relationship only holds for European options. If we consider American options, that can be exercised before the maturity date T, the value of the first portfolio might be different from the value of the second, and the parity breaks down.

1.3 Risk neutral pricing

Describe a market with possible paths described by a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$. Ω is the set of events that can occur. $\mathcal{F}$ is a sigma-algebra on Ω that contains all statements that can be be made about the events in Ω. The probability measure $\mathbb{P}$ (the 'real world measure') assigns probabilities to elements of $\mathcal{F}$. At last we define a filtration $\mathcal{F}_t$ on $\mathcal{F}$ that represents the flow of information. An price $S(t)$ then can be defined as a stochastic process $S(t, \omega)$ with $\omega \in \Omega$ adapted to the filtration $\mathcal{F}_t$.

A contingent claim (a contract) that pays out a single amount of cash at maturity time T can be represented by its payoff $H(\omega)$ for each event $\omega \in \Omega$. An example is a European call option that has payoff $(S_T - K)^+$ for some strike K. We wish to attribute a value to each contingent claim using the information that is known at each time: a *pricing rule*. Call this pricing rule $\Pi_t(H)$ for each contingent claim H. It should adhere to two constaints:

$$\forall \omega \in \Omega, H(\omega) \geq 0 \implies \forall t \in [0, T], \Pi_t(H) \geq 0, \tag{1.6}$$

$$\Pi_t \left(\sum_j H_j \right) = \sum_j \Pi_t(H_j). \tag{1.7}$$

The first constraint, positiveness, is necessary for we do not want to attribute a negative value to a claim that has a positive payoff for all events in Ω. The linearity constraint says that the sum of values of a set of contingent claims is equal to the value of a portfolio that contains those claims.

We will now get to the concept of an *arbitrage free pricing rule*. We can define the contingent claim $\mathbf{1}(\omega) = 1$ as the claim that pays out 1 unit of currency at time T in any event (a risk

free zero-coupon bond). We define the value of this bond as $\Pi_t(\mathbf{1}) = e^{-r(T-t)}$ with r the risk free interest rate. Now consider a mapping $\mathbb{Q} : \mathcal{F} \to \mathbb{R}$ defined as

$$\mathbb{Q}(A) \quad := \quad \frac{\Pi_0(1_A)}{\Pi_0(\mathbf{1})} = \Pi_0(1_A)e^{r(T-t)} \quad : A \in \mathcal{F}. \tag{1.8}$$

We can check that $\mathbb{Q}$ satisfies the requirements for a probability measure (if we allow countably infinite sums in the linearity constraint). We call this measure the *risk neutral measure*. Assuming additional continuity properties (Harrison and Pliska [1981]) on the set of contingent claims $\mathcal{H}$, we obtain the pricing rule

$$\Pi_0(H) \quad = \quad e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[H|\mathcal{F}_t] \tag{1.9}$$

which we call the *risk neutral pricing formula*. We have seen that any linear pricing rule is associated with a probability measure $\mathbb{Q}$. This measure does not necessarily represent real-world probabilities of events occurring; it just describes the prices of contingent claims.

We are now in a position to define the concept of an *arbitrage*. Loosely it is defined as an opportunity of a riskless profit: a trading strategy with zero initial investment that yields a positive payoff under a set that has a positive probability. We do not want these opportunities to exist in our model. Let us put this idea in a mathematical perspective.

Suppose there exists an event $A \in \Omega$ that occurs with zero probability: $\mathbb{P}(A) = 0$. Now define the contingent claim $\mathbf{1}_A$ that pays 1 if A occurs, and zero otherwise. The risk neutral pricing formula gives us a value for this claim:

$$\Pi_0(\mathbf{1}_A) = e^{-r(T-t)}\mathbb{Q}(A). \tag{1.10}$$

But this value must be zero as otherwise we would have an arbitrage opportunity by selling the contingent claim (as the event A will happen with probability zero). This implies that $\mathbb{Q}(A) = 0$. We can also show that if we assume $\mathbb{Q}(A) = 0$, then $\mathbb{P}(A) = 0$. Hence, the absence of arbitrage as we defined it here, implies that the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent (they assign probability zero to the same events).

At last we show that the discounted stock price is a martingale under $\mathbb{Q}$. Define $H = S_T$ to be the contingent claim that pays exactly the price of the stock S_T at time T, and use the risk neutral pricing formula to determine the price of this claim: $\Pi_0(S_T) = e^{-r(T-t)}\mathbb{E}[S_T|\mathcal{F}_t]$. But at time zero, this claim must be worth exactly S_t, as we can sell or buy the stock (and it has the same payoff at time T, regardless of the event ω). So we have found the relationship $S_t = e^{-r(T-t)}\mathbb{E}_{\mathbb{Q}}[S_T|\mathcal{F}_t]$ which implies that $e^{-rt}S_t$ is a martingale under $\mathbb{Q}$.

We conclude that an arbitrage free linear pricing rule is defined by a probability measure $\mathbb{Q}$ that is equivalent to $\mathbb{P}$ and under which the discounted stock price $e^{-rt}S_t$ is a martingale. The price of the contingent claim is then given by the risk neutral pricing formula (1.10).

1.3.1 Market completeness and the fundamental theorem of asset pricing

The existence and uniqueness of this risk neutral measure $\mathbb{Q}$ is the key to pricing contingent claims. It is dependent on the model we assume of the assets involved. When we assume the Black-Scholes market where stocks move as a diffusion, it was shown in Harrison and Pliska [1981] that this measure exists and is unique.

In a Lévy market where we assume the existence of jumps in the stock, this uniqueness (and as such the uniqueness of prices of contingent claims) is no longer guaranteed. It depends on the particular model if we can define a risk neutral measure (Cont and Tankov [2004], section 9.4). In the case of a model that has both positive and negative jumps, we can indeed define a risk neutral measure $\mathbb{Q}$, which is not guaranteed to be the unique measure. This implies that a unique price for a contingent claim is also not guaranteed. This problem is caused by the fact that jumps in the stock process cause a risk that cannot be hedged. It therefore depends on the preference of the observer to assign a price to it, that is not directly given by a no-arbitrage argument. This problem shall not be the topic of this thesis, but it is important to know when studying markets modeled by Lévy processes.

1.4 The Greeks and Delta Hedging

In options trading, the *Greeks* play an important role. They are defined as the partial derivatives of the option price to its variables. For the Black-Scholes model we have analytical solutions for the Greeks (for European calls). The simplest three are:

Name	Definition	Formula (Black-Scholes)
Delta	$\frac{\partial C}{\partial S}$	$N(d_1)$
Gamma	$\frac{\partial^2 C}{\partial S^2}$	$\frac{N'(d_1)}{S\sigma\sqrt{T-t}}$
Vega	$\frac{\partial C}{\partial \sigma}$	$SN'(d_1)\sqrt{T-t}$

where $N(x)$ is the normal distribution function and $d_1 = \frac{\log(S/K)+(r+\frac{\sigma^2}{2})T}{\sigma\sqrt{T}}$. The delta represents the velocity at which the option price changes with respect to a change in the stock price. The gamma measures the sensitivity of the delta with respect to the stock price, and the vega measures the sensitivity of the option price with respect to the volatility parameter. Of course these Greeks are entirely dependent on the model that one assumes. Option market makers monitor these variables closely. They *make a market* in options by providing *bid and ask prices* for which they can buy and sell options. There is a difference between these prices called the *spread* that is a source of income for the market maker. The market maker usually engages in a portfolio of different assets (such as stocks or futures) to eliminate the risks that are associated to taking an options position. *Delta hedging* then refers to keeping the *delta* of the portfolio (the derivative of the portfolio with respect to the stock price) equal to zero.

1.5 Implied Volatility

If we observe market prices of options, we might want to see what variables of the Black-Scholes model coincide with market prices. As we usually know four of the five variables needed - the stock price S, the strike K, the interest rate r and the time to maturity $T-t$, we are left with a function of the volatility σ. If a market price of some European call option is c, we can approximate the *implied volatility* σ_{imp} by finding the solution of

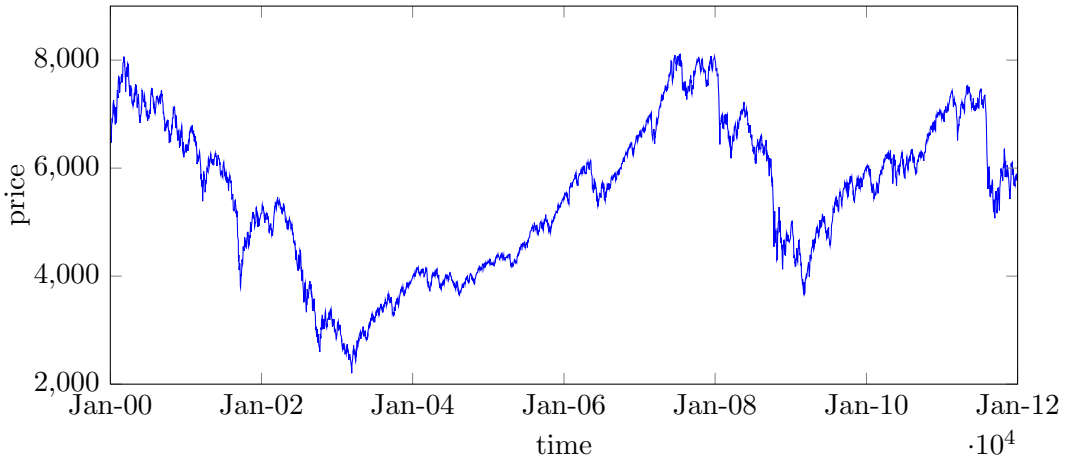
$$C(\sigma) = c.$$

This is possible since C is increasing in σ (as the *vega* is always larger than zero), and we can use for instance Newton's method to approximate it.

1.6 Shortcomings of the Black-Scholes model

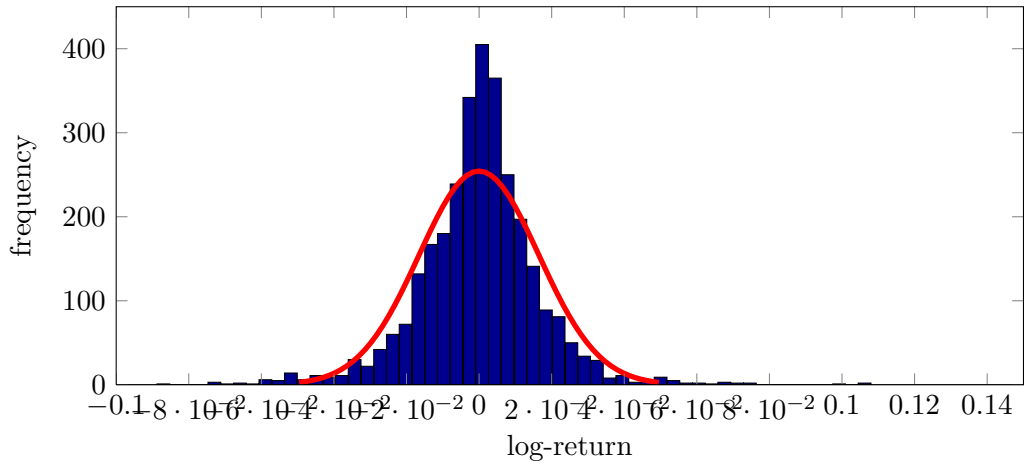
Our motivation from using a Lévy process as a model for an equity index comes from some shortcomings of the Black-Scholes model. We will take a look at some obvious ones. Let us have a look at the path of the DAX index in the period 2000-2011 and model this index by a stochastic process S_t:

Figure 1.1: DAX index during 2000-2011



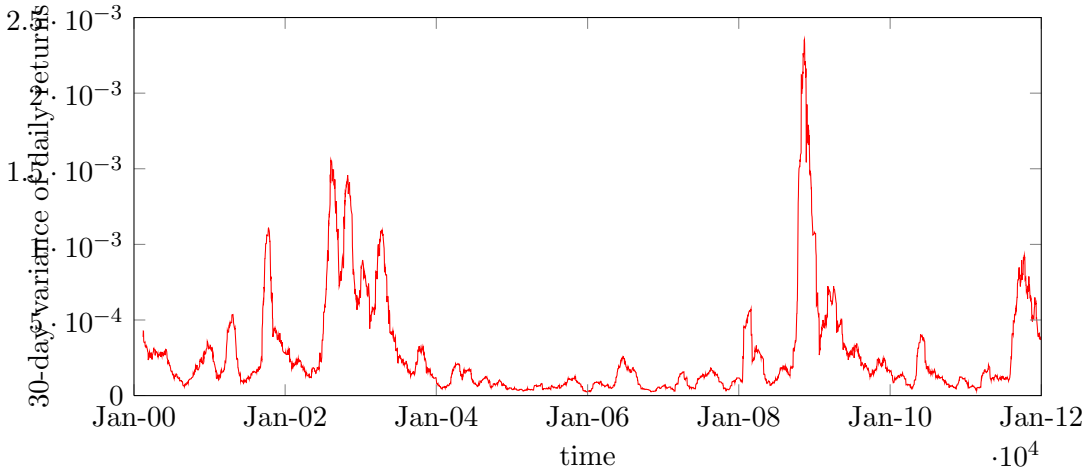
Within the Black-Scholes model we would model the log-returns $\log(\frac{S_{t+1}}{S_t})$ by a normal distribution. If we make a Q-plot of the log-returns, we obtain:

Figure 1.2: Daily log-returns of the DAX index in the period 2000-2011 versus a fitted normal distribution



This clearly shows the non-normality of the returns: the peak is much sharper and the tails are fatter than the normal distribution. Now take the 30-day variance of the log-returns. In the Black-Scholes model, this variance should be approximately constant.

Figure 1.3: 30-day daily variance of the DAX index during 2000-2011

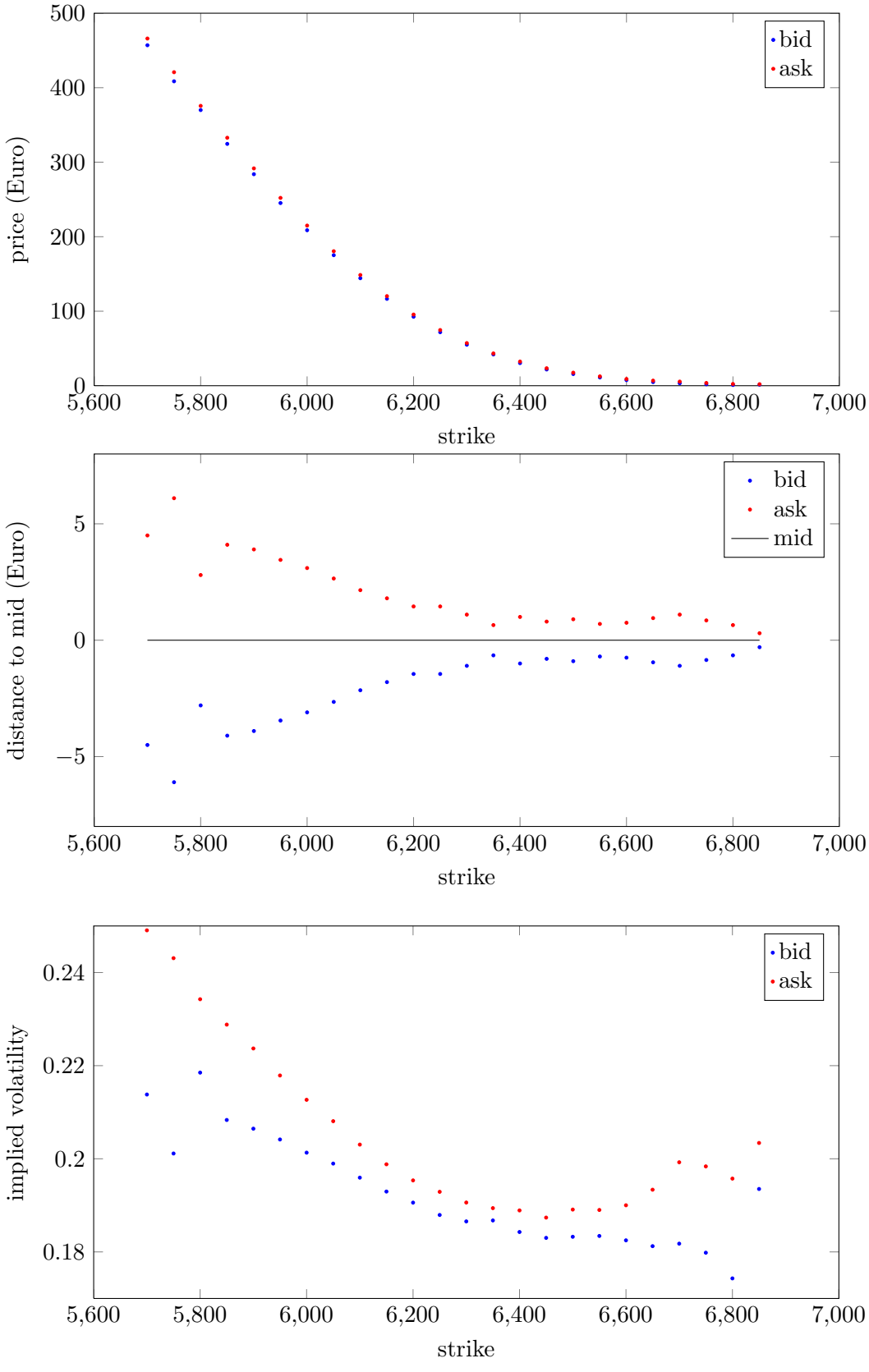


We see that first of all this variance is far from constant. We notice a second thing: a dependence structure of the variance over time. If the variance is high at some point in time, it is likely to still be high at the next point in time, and so in. This *volatility clustering* is important in modeling the price of the stock index. In this thesis we will show that a Lévy process can solve the problem of the distribution of the log-returns of the process. The problem of volatility clustering cannot be solved by a Lévy process alone.

1.7 Representations of prices throughout this thesis

There are multiple ways to view option prices. The most simple way is of course to show the actual prices (see Figure 1.4). However this view can prohibit us from seeing fine details of the prices, as the small strikes (the *in the money* options) have a much larger scale than the large strikes (the *out of the money* options. Another way to show the prices is to view the bid and ask prices as distances to the mid price. This is shown in the middle graph of Figure 1.4. A desirable property is that this view is model independent. A drawback is that the mid price does not necessarily reflect a 'true' price. At last we can also display the prices as *implied (Black-Scholes) volatility*, as explained in a previous section. A desirable feature is that it shows what we call the volatility smile (the strike-dependency of the implied volatility, a feature that shows that the Black-Scholes normality assumption is inaccurate). If the Black-Scholes model were a true description of the market, this should be a straight line. Therefore it shows how the market deviates from this (log)normality assumption.

Figure 1.4: DAX Jan-12 Call on Jan-04-2012, 11:15



To construct the implied volatility graph, we need to assume an interest rate r that corresponds to this specific maturity. In the above picture we have taken $r = 0.01$ as a constant. However throughout this thesis we keep the interest rates as free parameters and calibrate these to market data of call and put prices.

1.8 Lévy processes

A stochastic process X_t is a Lévy process if it has independent and stationary increments, and it is stochastically continuous (for any $\epsilon > 0, \lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| > \epsilon) = 0$). Two examples of a Lévy process are the Poisson process and Brownian motion.

For a Poisson process it is easy to show that is a Lévy process. It is defined as the counting process N_t with $N_0 = 0$, that has independent and stationary increments, and the increments have probabilities defined by $\mathbb{P}(N_{t+h} - N_t = n) = e^{-\lambda h}\frac{(\lambda h)^n}{n!}$. We only need to prove the stochastic continuity. Considering the fact that it is an increasing process, we have $\mathbb{P}(|N_{t+h} - N_t| > \epsilon) = \mathbb{P}(N_{t+h} - N_t > \epsilon) > \mathbb{P}(N_{t+h} - N_t > 0) = 1 - \mathbb{P}(N_{t+h} - N_t = 0) = 1 - e^{-\lambda h}$. This tends to zero as $h \to 0$, which proves the stochastic continuity property.

A Lévy process is intimately connected to the notion of infinitely divisible distributions.

Definition 2. *Infinitely divisible distribution Consider a random variable X with cumulative distribution function F. We call X infinitely divisible if and only if for every $n = 1, 2, \ldots$ we can decompose X as*

$$X = X_1 + X_2 + \ldots + X_n \tag{1.11}$$

for some independent and identically distributed random variables X_i.

The increments of a Lévy process are infinitely divisible; we can argue that an increment over the interval $[t, t + h]$ is a sum of n increments. From the independence and stationarity properties of the Lévy process, it then follows that these n increments are independent and identically distributed. Therefore each increment of the Lévy process is infinitely divisible. This leads to the thought that a Lévy process is a continuous time generalization of a sum of independently and identically distributed random variables.

The pure jump perspective

We could wonder what is the best way to build up a Lévy model with the three ingredients drift, diffusion and jumps. The drift is easily determined. As in the risk neutral framework, we define the stock price S_t to be $S_t = \exp(L_t + rt)$, with L_t a Lévy process that is a martingale. We can therefore safely assume the drift to be zero. The diffusion component is a more subjective topic. One can think of the diffusion component of describing the 'smooth' path behaviour of an asset, while jumps describe swift changes that can occur in the market.

In this thesis we will choose the perspective where we see all behaviour in the market as jumps. Since we observe the market as a discrete phenomenon, we could argue that a jump process is enough to describe it. At last we also assume the Lévy jump measure ν to have a density $\nu(x)$.

Lévy Khintchine theorem

A key result in the theory of Lévy processes is the Lévy-Khintchine theorem, that describes the connection between the characteristic function of the process with its drift, diffusion and Lévy

measure components. Its general version (for a real-valued process) reads

$$\phi_t(u) = \mathbb{E}[e^{iuL_t}] \quad = \quad \exp\left\{t\left(i\gamma u - \frac{1}{2}\sigma u^2 + \int_{\mathbb{R}}(e^{iux} - 1 - iux1_{|x|\leq 1})\nu(dx)\right)\right\}$$

Where γ is a drift constant, σ is the volatility of the diffusion component, and ν is the Lévy measure. In the finite variation case (a diffusion is absent and $\int_{|x|\leq 1}|x|\nu(dx) < \infty$) it simplifies to (Cont and Tankov [2004], Corollary 3.1)

$$\phi_t(u) = \mathbb{E}[e^{iuL_t}] \quad = \quad \exp\left\{t\left(iub + \int_{\mathbb{R}}(e^{iux} - 1)\nu(dx)\right)\right\}.$$

It is important to understand that the definition of γ, σ, ν fixes the whole stochastic process L_t. What we will do in practice is define the one-year characteristic function $\phi_1(u) = \mathbb{E}[e^{iuX_1}]$ and use the relationship $\phi_t(u) = \phi_1(u)^t$ (see section 2.2 for a proof) to define the characteristic function at other points in time.

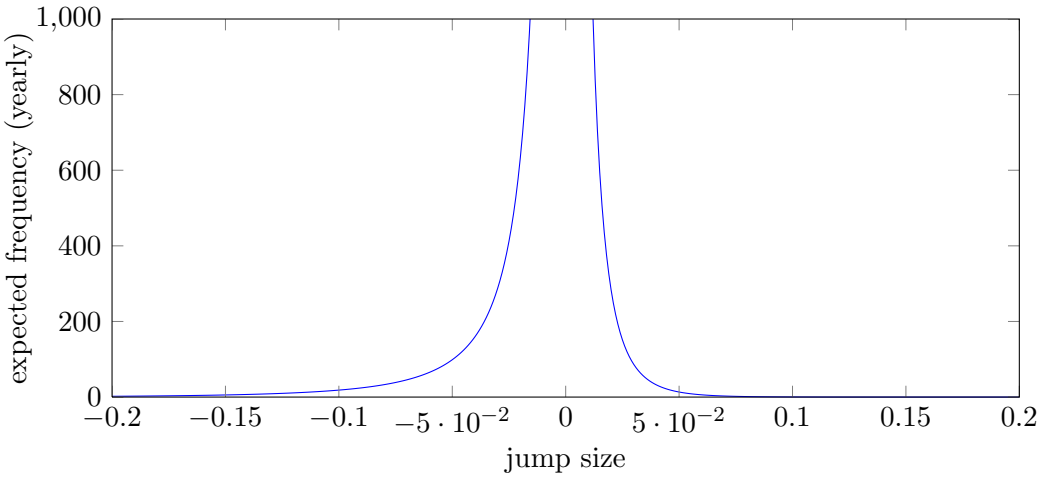
The Lévy density

If we now assume the Lévy measure ν to have a density $\nu(x)$, we can try to find a parametric model to fit our data. An example is the CGMY model (Carr et al. [2002]), that defines the Lévy density as

$$\nu_{\text{CGMY}}(x) \quad = \quad \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}} & \text{if } x < 0 \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}} & \text{if } x > 0 \end{cases}.$$

For the CGMY process, the definition adheres to the constraint $\int \min(1, x^2)\nu(x)dx < \infty$ only if $Y < 1$. This condition is necessary for the finite variation case of the Lévy-Khintchine formula.

Figure 1.5: Example of the CGMY Lévy density



A graph of this density with parameters $C = 1, G = 10, M = 40, Y = 0.7$ shows the singularity at zero. The meaning of the constraint $\int \min(1, x^2)\nu(x)dx < \infty$ is that an infinite amount of jumps can occur for small jumps ($|x| \leq 1$), but not for large jumps. For small jumps the variance should be finite. The practical meaning of this Lévy density: $\int_a^b \nu(x)dx$ *is the expected number, per unit time, of jumps whose size belong to* $\int_a^b \nu(x)dx$.

1.9 The Augmented CGMY process and an outline of the thesis

Our project is about finding a good parametric representation of the data. The Augmented CGMY process acknowledges that the CGMY model does not fit well in the tails of the density (we will show this in the thesis), and as such modifies the tails of the CGMY process. Its Lévy density is defined as

$$\nu_{\text{Augmented CGMY}}(x) = \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}} & \text{if } \theta_l < x < 0 \\ C\frac{\exp(-G|x|)\exp(-L|x-\theta|)}{|x|^{1+Y}} & \text{if } x \leq \theta_l \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}} & \text{if } 0 < x < \theta \\ C\frac{\exp(-M|x|)\exp(-R|x-\theta|)}{|x|^{1+Y}} & \text{if } x \geq \theta_r \end{cases}.$$

To illustrate the different behaviour of the two models, we show a slice of option data with the CGMY model and the Augmented CGMY model calibrated to it.

Figure 1.6: DAX Nov-11 Call on Nov-15-2011, 11:15 (CGMY)

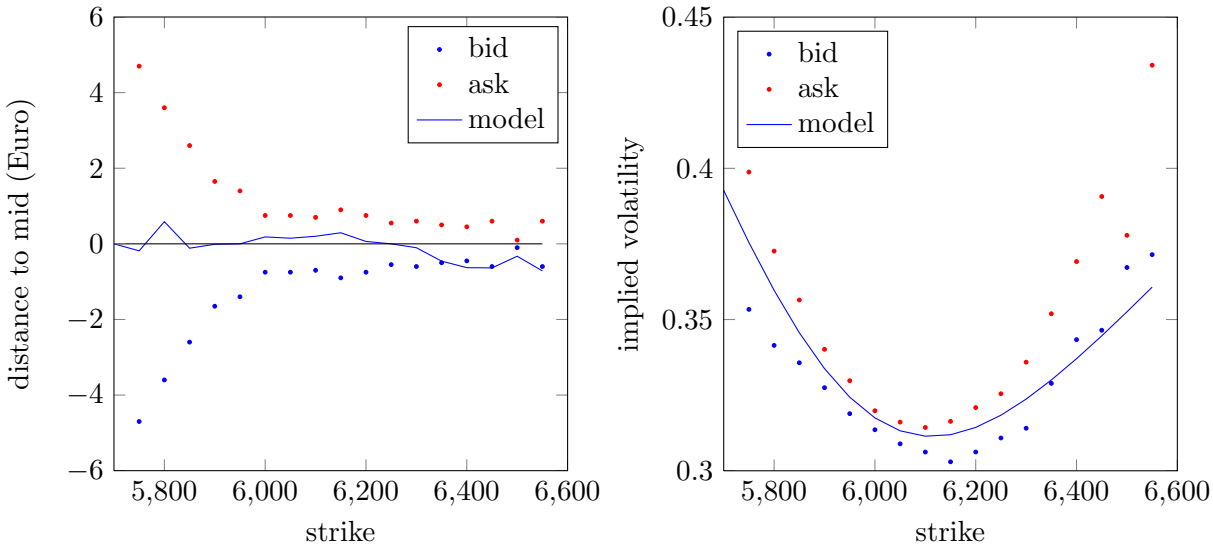
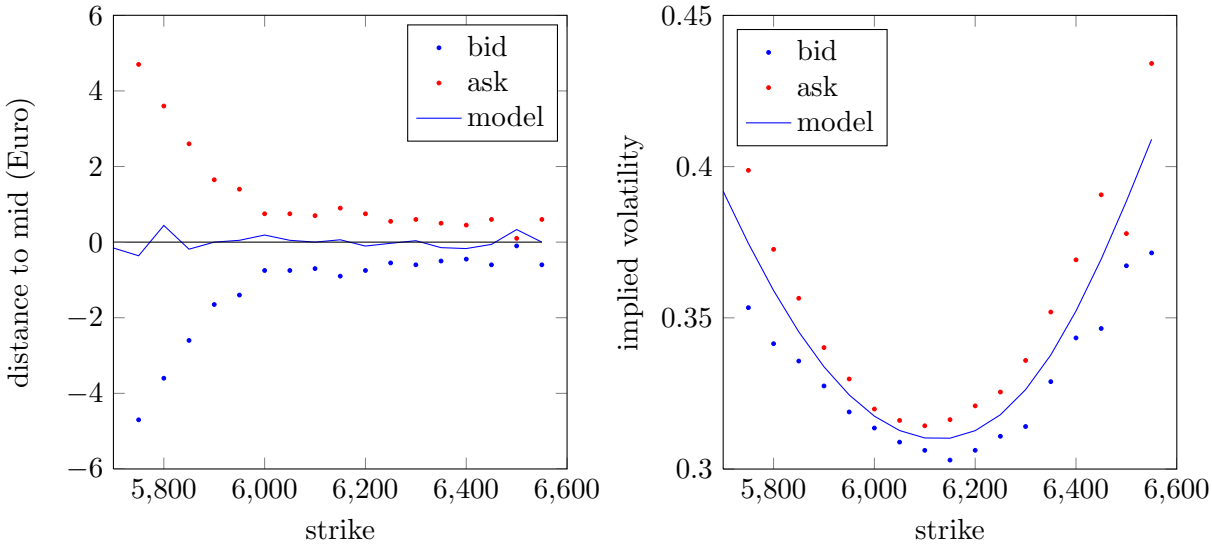


Figure 1.7: DAX Nov-11 Call on Nov-15-2011, 11:15 (Augmented CGMY)



The option class shown in the graph above is the November Call option that expired on 18 November 2012 at 13:00 CET. At the time of the slice it has just over three days left until maturity. In this particular slice the CGMY model does not fit well for the higher strikes. It has a bias to underprice the out of the money options. In the second graph we see the Augmented CGMY model with 8 parameters fit to the same slice. We can see that it fits better in that region of the slice. It is not a perfect fit: we see that the second-largest maturity does not fit. However notice that the ask price is significantly lower than the adjacent ask prices. This might be a price that we do not want our model to fit, as it might not represent the 'real' price.
Of course the above slice is just one slice of data. In the thesis we will show a more extensive statistical analysis of the performance of different Lévy models, including a few variants of the Augmented CGMY model. We discuss additional problems (such as the weighting of observations and the loss function) that we face in the calibration of the model to the data. In particular we will reach the following conclusions.

The Augmented CGMY process fits the option prices significantly better than the CGMY process, which itself fits the best of the evaluated existing models (Black Scholes, Variance Gamma, NIG, Meixner, CGMY). Lévy processes do not fit all the option maturities simultaneously. This is due to the assumptions of the Lévy process that are too restrictive.

Chapter 2

The Lévy framework

We will define the Lévy process and summarize its important properties. Cont and Tankov [2004] serves as our guide through the theory. For more advanced proofs we shall refer to Sato [1999]. Our approach is as following. First we introduce the main definitions and the main theorems (the Lévy-Itô decomposition and the Lévy-Khintchine formula). We introduce simple properties of the Lévy process such as its moments and path properties. At last we define our market model (the exponential Lévy process) and describe two measure transforms that allow us to move between the real-world measure $\mathbb{P}$ and the risk neutral measure $\mathbb{Q}$.

2.1 Characteristic functions and Fourier transforms

In the theory of Lévy processes it is natural (following from the shortly introduced Lévy-Khintchine representation) to work with the *characteristic function* of a density as opposed to the density itself. A *characteristic function* $\phi(u)$ of a random variable X with respect to some probability measure $\mathbb{P}$ is defined as

$$\phi(u) := \mathbb{E}_{\mathbb{P}}[e^{iuX}] = \int e^{iux} f(x) dx \tag{2.1}$$

where the right hand side part is true if the probability density function $f(x)$ of X exists. We see that the characteristic function is similar to what we usually refer to as the inverse Fourier transform. A good introduction to the theory of characteristic function is Lukacs [1960]. We will discuss some of its properties that we will use throughout the theory.

First of all, note that $\phi(0) = 1$ holds for any distribution. Suppose that we know the characteristic function $\phi(u)$. We can then invert this to obtain the distribution using the Fourier transform:

$$f(x) \quad = \quad \frac{1}{2\pi} \int e^{-iux} \phi(u) du. \tag{2.2}$$

What we will encounter often are translated distributions such as $g(x) := f(x - a)$ for some constant $a \in \mathbb{R}$. Its characteristic function is given by the translation identity:

$$\phi_g(u) = e^{iua} \phi_f(u) \tag{2.3}$$

where ϕ_g is the characteristic function of g and ϕ_f that of f. If we have a sum of independent random variables $Z = X + Y$, then the characteristic function of Z is

$$\phi_Z(u) \quad = \quad \phi_X(u) \phi_Y(u). \tag{2.4}$$

Another convenient property of characteristic function is the fact that we can obtain the moments of the random variable (assuming they exist) using the derivatives of the characteristic function evaluated at zero:

$$\mathbb{E}[X^n] = \frac{1}{i^n}\phi^{(n)}(0). \tag{2.5}$$

Proof of this is found in Lukacs [1960] section 1.4.

2.2 Definition of the Lévy process

Definition 3. *A stochastic process $(X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ such that $X_0 = 0$ is a Lévy process if it has*

1. *Independent increments: for every increasing sequence of times $t_0, \ldots, t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*
2. *Stationary increments: the distribution of $X_{t+h} - X_t$ does not depend on t.*
3. *Stochastic continuity: for all $\epsilon > 0$, $\lim_{h\to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \epsilon) = 0$.*

Usually we assume that the process is càdlàg. A càdlàg stochastic process ("continue à droite, limite à gauche") is right continuous and has left limits. It is proved in Protter [2004] that every Lévy process has a unique càdlàg modification. For the modeling of financial instruments, the càdlàg property is a natural way of representing the flow of information through time.

The continuity property does not imply that the sample paths are continuous at all. However it does guarantee that the process does not exhibit jumps (discontinuities) at fixed (nonrandom) times. In the context of financial modeling this might be an issue: one might anticipate a jump (of known or unknown size) in the stock at some fixed time - such as an earnings announcement. This cannot be modeled by a Lévy process.

Two basic examples of Lévy processes are the Poisson process and the Brownian motion. In this sense it has two attractive properties for modeling financial assets: it allows for both jumps (the Poisson process) and a diffusion component (the Brownian motion). As a sum of two Lévy processes is again a Lévy process, it is possible to construct such a process.

The concept of infinite divisibility is closely related to Lévy processes:

Definition 4 (Infinite divisibility)**.** *A probability distribution F on $\mathbb{R}^d$ is said to be infinitely divisible if for any integer $n \geq 2$, there exist n i.i.d. random variables $Y_1, \ldots, Y_n$ such that $Y_1 + \ldots + Y_n$ has distribution F.*

A well-known result (see for example Sato [1999], Theorem 7.10) is that Lévy processes and infinitely divisible distributions are related:

Theorem 1. *A Lévy process $(X_t)_{t\geq 0}$ has an infinitely divisible distribution for every t. Conversely, if F is an infinitely divisible distribution, then there exists a Lévy process $(X_t)_{t\geq 0}$ such that the distribution of X_1 is given by F.*

Now take any two positive integers m,n and apply the infinite divisibility property twice:

$$X_m = \underbrace{X_1 + (X_2 - X_1) + \ldots + (X_m - X_{m-1})}_{m \text{ random variables}} \tag{2.6}$$

$$X_m = \underbrace{X_{\frac{m}{n}} + (X_{2\frac{m}{n}} - X_{\frac{m}{n}}) + \ldots + (X_m - X_{(n-1)\frac{m}{n}})}_{n \text{ random variables}}. \tag{2.7}$$

Since these are sums of independent identically distributed variables, we can calculate the characteristic functions as

$$\begin{aligned} \phi_m(u) &= \phi_1(u)^m && (2.8)\\ \phi_m(u) &= \phi_{\frac{m}{n}}(u)^n. && (2.9) \end{aligned}$$

Equating the two gives $\phi_{\frac{m}{n}}(u) = (\phi_1(u))^{\frac{m}{n}}$, and this implies that for any rational $t = \frac{m}{n}$

$$\phi_t(u) = \phi_1(u)^t. \quad (2.10)$$

Actually, using Lévy's continuity theorem we can show that for any $t \in \mathbb{R}$ the result holds. Since the set of rationals is dense in $\mathbb{R}$, for any irrational number t we can find a sequence of rationals $(t_n)_{n\in\mathbb{N}}$ such that $t_n \to t$. Now construct the sequence of random variables $(X_{t_n})_{n\in\mathbb{N}}$ from some Lévy process X. From the convergence of the sequence (t_n) it follows that (X_{t_n}) converges in distribution to X_t. Lévy's continuity theorem states that for any sequence of random variables that converge in distribution, their characteristic functions $(\phi_{t_n}(u))_{n\in\mathbb{N}}$ converge pointwise in u (to $\phi_t(u)$). Now we are in a position to show the result. For N large, we have

$$\begin{aligned} |\phi(u)^t - \phi_t(u)| &\leq |\phi(u)^t - \phi_{t_N}(u)| + |\phi_{t_N}(u) - \phi_t(u)| \\ &= |\phi(u)^t - \phi_{t_N}(u)| + \epsilon_1 \\ &= |(\phi(u)^{t-t_n} - 1)\phi(u)^{t_N}| + \epsilon_1 \\ &\leq L|(\phi(u)^{t-t_n} - 1)| + \epsilon_1 \\ &= L|(\phi(u)^{\epsilon_2} - 1)| + \epsilon_1 \end{aligned}$$

where L is constant and ϵ_1, ϵ_2 are arbitrarily small. The right hand side vanishes as $\epsilon_1, \epsilon_2 \to 0$, so we conclude that $\phi_t(u) = \phi(u)^t$ for any $t \in \mathbb{R}$.

Now assuming that we can write any characteristic function $\phi(u)$ as $\phi(u) = \exp(f(u))$ for some continuous function $f : \mathbb{R} \to \mathbb{C}$ (this is proven in Sato [1999], Theorem 7.6), we have proven the following.

Theorem 2 (Characteristic exponent)**.** *For every Lévy process $(X_t)_{t\geq 0}$, there exists a continuous mapping $\psi : \mathbb{C} \to \mathbb{C}$ called the characteristic exponent of X, such that*

$$\mathbb{E}[e^{iuX_t}] = \exp(t\psi(iu)) \quad \text{where } u \in \mathbb{C}.$$

What is important to realize from this definition of the Lévy process, is that we can fix the distribution of the process at one point in time (it is customary to fix it at time one), and that the process is then fixed through time. We will now state some important results from the works of Lévy, Itô and Khintchine.

2.2.1 The Lévy-Itô decomposition

The main result of the theory is the decomposition of the Lévy process in three different parts: a drift, a diffusion (or Brownian motion) and a compensated jump part. We first the definition of a Poisson random measure and a Lévy measure. We follow Cont and Tankov [2004] closely.

Definition 5 (Poisson random measure)**.** *We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with μ a positive Radon measure on $(E, \mathcal{E})$, where $E \subset \mathbb{R}^+$, and $\mathcal{E}$ the σ-algebra of Borel sets of E. A Poisson random measure on E with intensity measure μ is a function $M : \Omega \times \mathcal{E} \to \mathbb{N}$ such that*

1. *For almost all $\omega \in \Omega$, $M(\omega, .)$ is an integer-valued Radon measure on E.*

2. *For each measurable set $A \subset E$, $M(.,A) = M(A)$ is a Poisson distributed random variable with intensity $\mu(A)$.*

3. *For disjoint measurable sets $A_1, \ldots, A_n \in \mathcal{E}$, the variables $M(A_1), \ldots, M(A_n)$ are independent.*

The Poisson random measure on E can be seen as a random variable taking values in the set of Radon measures on E. It controls the frequency of jumps occurring. We now introduce the Lévy measure, that controls the size of the jumps. The Lévy measure $\nu(A)$ is defined as the expected number of jumps with size in A per unit time:

Definition 6 (Lévy measure)**.** *If $(X_t)_{t\geq 0}$ is a Lévy process on $\mathbb{R}^d$, then the Lévy measure ν on $\mathbb{R}^d$ is defined by*

$$\nu(A) = \mathbb{E}[\#\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}], \quad \text{where } A \in \mathcal{B}(\mathbb{R}^d).$$

Theorem 3 (Lévy-Itô decomposition)**.** *If $(X_t)_{t\geq 0}$ is a Lévy process, and ν is its Lévy measure, then*

- *ν is a Radon measure on $\mathbb{R} \setminus \{0\}$ and*
$$\int |x|^2 \nu(dx) < \infty.$$
- *The jump measure of X, denoted by J_X is a Poisson random measure on $[0,\infty) \times \mathbb{R}$ with intensity measure $\nu(dx)dt$.*
- *There exists a vector γ and a Brownian motion $(B_t)_{t\geq 0}$ with variance σ^2 such that*
$$\begin{aligned} X_t &= \gamma t + B_t + X_t^l + \lim_{\epsilon \downarrow 0} \widetilde{X}_t^\epsilon \\ X_t^l &= \int_{|x|\geq 1, s\in[0,t]} x J_X(ds \times dx) \\ \widetilde{X}_t^\epsilon &= \int_{\epsilon \leq |x| < 1, s\in[0,t]} x\{J_X(ds \times dx) - \nu(dx)ds\} \\ &= \int_{\epsilon \leq |x| < 1, s\in[0,t]} x \widetilde{J}_X(ds \times dx) \end{aligned}$$

where the terms are independent and convergence is almost sure and uniform in t on $[0,T]$.

This means that any Lévy process can be characterized as a *characteristic triplet* (σ^2, ν, γ), where σ is the variance of the Brownian motion, γ is a drift parameter, and ν is a Lévy measure that satisties $\int |x|^2 \nu(dx) < \infty$. A proof can be found in Sato [1999], Theorem 19.2, 19.3.

2.2.2 The Lévy-Khintchine representation

The independence of the added components of drift, diffusion and compensated jumps allow an analytical format for the characteristic function of the Lévy process, the Lévy-Khintchine representation:

Theorem 4 (Lévy-Khintchine representation)**.** *If $(X_t)_{t\geq 0}$ is a Lévy process on $\mathbb{R}$ with characteristic triplet (σ^2, ν, γ), then*

$$\begin{aligned} \phi_{X_t}(u) &= \mathbb{E}[e^{iuX_t}] = e^{t\psi(iu)} \\ \text{with} \quad \psi(u) &= -\frac{1}{2}\sigma u^2 + i\gamma u + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1 - iux 1_{|x|\leq 1})\nu(dx). \end{aligned}$$

In the finite variation case, this simplifies to

$$\begin{aligned} \phi_{X_t}(u) &= \mathbb{E}[e^{iuX_t}] = e^{t\psi(iu)} \\ \text{with} \quad \psi(u) &= ibu + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1)\nu(dx). \end{aligned}$$

as $b = \gamma - \int_{|x|\leq 1} x\nu(dx) < \infty$, and the diffusion component (which is of infinite variation) is absent. In the rest of this thesis we shall use jump only processes with finite variation and thus use this version of the Lévy-Khintchine representation.

2.3 Moments and cumulants of the Lévy process

2.3.1 Moments and cumulants

Moments of a probability distribution describe the shape of the distribution. The *raw moments* μ_n are defined as

$$\mu_n = \mathbb{E}[X^n] = \int x^n f(x)dx. \tag{2.11}$$

The *central moments* are defined such that they are translation invariant:

$$\mu'_n = \mathbb{E}[(X - \mathbb{E}[X])^n] = \int (x - \mathbb{E}[X])^n f(x)dx. \tag{2.12}$$

Moments of a distribution do not necessarily exist. Just like the moments describe the properties of a distribution, we can use the *cumulants* κ_n of the distribution. First define the *cumulant generating function* $g(z)$:

$$g(z) := \log\left(\mathbb{E}[e^{zX}]\right) \quad : z \in \mathbb{R}. \tag{2.13}$$

Or in terms of the characteristic function ϕ and the characteristic exponent ψ:

$$g(z) = \log\left(\phi(-iz)\right) = t\psi(z). \tag{2.14}$$

The cumulants κ_n are then defined as the derivatives of the cumulant generating function, evaluated at zero:

$$\kappa_n := g^{(n)}(0) = t\psi^{(n)}(0). \tag{2.15}$$

Cumulants and moments are interchangable. That is, if one knows the moments of a distribution, one can calculate the cumulants, and vice versa. To calculate the first four central moments from the first four cumulants, we can use the following formulas.

$$\begin{aligned} \mu'_1 &= \kappa_1 \\ \mu'_2 &= \kappa_2 + \kappa_1^2 \\ \mu'_3 &= \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3 \\ \mu'_4 &= \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4 \end{aligned} \tag{2.16}$$

Or the other way around from the moments to the cumulants.

$$\begin{aligned} \kappa_1 &= \mu'_1 \\ \kappa_2 &= \mu'_2 - {\mu'_1}^2 \\ \kappa_3 &= \mu'_3 - 3\mu'_2\mu'_1 + 2{\mu'_1}^3 \\ \kappa_4 &= \mu'_4 + 4\mu'_3\mu'_1 - 3{\mu'_2}^2 + 12\mu'_2{\mu'_1}^2 - 6{\mu'_1}^4 \end{aligned} \tag{2.17}$$

From a financial perspective the moments are interesting since they describe a distribution in a neat way: the mean, variance or skewness of a distribution give us a quick indication of its characteristics. In the setting of Lévy processes the cumulants are useful as they are extracted directly from the characteristic exponential.

2.4 Moments of the exponential Lévy process

As a means to describe the properties of the distribution of the stock price modeled by an exponential Lévy process $S_t = \exp(L_t)$, we are interested in its raw moments $\mathbb{E}[S_t^n]$. Any central moments or cumulants of order n follow from these if they exist. As in this thesis we are mainly concerned with the properties of the process under the risk neutral measure $\mathbb{Q}$, we assume the expectations to be under $\mathbb{Q}$. The raw moments are easily obtained via the characteristic function of the Lévy process L_t:

$$\mathbb{E}[S_t^n] \quad = \quad \mathbb{E}[e^{nL_t}] = \mathbb{E}[e^{i(-in)L_t}] = \phi_{L_t}(-in)$$

However, it is not guaranteed at all that these moments exist. If we assume a pure jump process L_t, we can write the moments in terms of the Lévy density ν:

$$\mathbb{E}[S_t^n] \quad = \quad \exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{nx}-1)\nu(x)dx\right)$$

we see that the nth moment only exists if

$$\int_0^\infty e^{nx}\nu(x)dx < \infty$$

For the CGMY process this holds if $M > n$, which in practice is easily violated. Therefore we can also consider the moments of the log-stock price process, the Lévy process L_t. We can calculate these through the cumulants κ_n (see Cont and Tankov [2004], Proposition 3.13), that are calculated for a pure jump Lévy process by

$$\kappa_n \quad = \quad t\int_{\mathbb{R}\setminus\{0\}} x^n\nu(x)dx$$

The existence of these moments require only polynomial decay in the tails, and thus in the case of CGMY exist for all orders as the exponential decay of CGMY dominates polynomial decay of all orders.

2.4.1 The case of a jump-only Lévy process

Assume that L_t is a pure jump Lévy process with characteristic function in Lévy-Khintchine form

$$\phi(u) \quad = \quad \exp\left(t\psi(iu)\right) \tag{2.18}$$

with the characteristic exponent ψ in terms of the Lévy density ν

$$\psi(z) \quad = \quad \int_{\mathbb{R}\setminus\{0\}}(e^{zx}-1)\nu(x)dx. \tag{2.19}$$

Now the derivatives of the characteristic exponential are easily determined and the cumulants of the distribution are:

$$\kappa_n \quad = \quad t\psi^{(n)}(0) = t\int x^n\nu(x)dx. \tag{2.20}$$

Although the Lévy density is not a density in the probabilistic sense (its integral is not equal to one), we shall refer to the above integrals as the 'moments' of the Lévy density.

The central moments can be calculated using equation (2.16). We observe that the cumulants of the Lévy process increase linearly over time. This is a property that follows from the infinite divisibility assumption.

2.5 Path properties

2.5.1 Variation and activity

Let $P = \{a = t_1 < t_2 < \cdots < t_{n-1} < t_n = b$ be a partition of an interval $[a, b]$, and $f(x)$ a real function. Then the variation of f over P is defined as

$$\text{var}_P(f) \quad = \quad \sum_{i=1}^{n} |f(t_{i+1}) - f(t_i)|.$$

Define the variation of f as the supremum over all possible partitions

$$\text{var}(f) \quad = \quad \sup_P \sum_{i=1}^{n} |f(t_{i+1}) - f(t_i)|.$$

Now if the variation is finite on all compact intervals, we call the function f of *finite variation*. Cont and Tankov [2004] section 3.5 proves the following result that connects variation with the Lévy density ν.

Theorem 5 (Finite variation Lévy process)**.** *A Lévy process is of finite variation if and only if it has no diffusion component and*

$$\int_{-1}^{1} |x|\nu(x)dx < \infty. \tag{2.21}$$

The necessary absence of the diffusion component is of course trivial as the Brownian motion is a process of infinite variation.

The activity of a Lévy process measures how many jumps are expected in a finite time interval. A Lévy process is of *finite activity* if

$$\int \nu(x)dx < \infty.$$

We shall encounter Lévy processes of both the finite and the infinite kind.

2.6 The assumed market model

We model the stock price as

$$S_t \quad = \quad S_0 \exp(X_t) \tag{2.22}$$

where X_t is a Lévy process. Hence we model the log-returns of the stock price as a Lévy process. In the thesis we shall be concerned with the logarithm of the stock price, as defined by

$$X_t := \log(S_t) = \log(S_0)L_t. \tag{2.23}$$

The characteristic function of the log-stock price process X_t is found via the translation property of the Fourier transform:

$$\phi_{X_t}(u) = \mathbb{E}[e^{iuX_t}] = e^{i\log(S_0)u}\phi_{L_t}(u) = (S_0)^{iu}\phi_{L_t}(u) \tag{2.24}$$

with ϕ_{L_t} the characteristic function of the Lévy process at time t.

2.7 Change of measure

In the introduction we have described the concept of arbitrage free pricing. It relies on finding a probability measure $\mathbb{Q}$ that has the properties

(i) $\mathbb{Q} \sim \mathbb{P}$

(ii) $\hat{S}_t = D_t S_t$ is a martingale under $\mathbb{Q}$

where D_t is a *discount process*, usually $D_t = e^{-rt}$ in the case of a constant interest rate r. Next to this it is a desirable property that the change of measure preserves our model. That is, if the process S_t is a Lévy process under $\mathbb{P}$, we would like it to also be a Lévy process under $\mathbb{Q}$. Even more desirable would be if our process would be in a certain class of models (CGMY, for instance) under $\mathbb{P}$, and is still in the CGMY class (with different parameters) under $\mathbb{Q}$. First of all it is not guaranteed that such a measure exists or is unique. To illustrate this concept, let us first show how this measure change could work for the Poisson process. For this we need the Radon-Nikodym theorem:

Theorem 6 (Radon-Nikodym)**.** *If $\mathbb{Q} \sim \mathbb{P}$, then there exists an almost surely positive random variable $\frac{d\mathbb{Q}}{d\mathbb{P}}$, called the Radon-Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$. For any random variable Z then*

$$\mathbb{E}_{\mathbb{Q}}[Z] \quad = \quad \mathbb{E}_{\mathbb{P}}\left[Z\frac{d\mathbb{Q}}{d\mathbb{P}}\right]. \tag{2.25}$$

The Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ describes the measure change, as we can deduce the probabilities of $\mathbb{Q}$ from the above result.

Example: measure change of the Poisson process

Suppose N_t is a homogeneous Poisson process with intensity parameter λ_1 under $\mathbb{P}$. We are interested to find a measure $\mathbb{Q} \sim \mathbb{P}$ such that N_t is a Poisson process with intensity λ_2. Using Radon-Nikodym we need to show that there exists a Radon-Nikodym derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ that leads to a Poisson process under both $\mathbb{P}$ and $\mathbb{Q}$. The Poisson process defines probabilities under $\mathbb{P}$ given by

$$\mathbb{P}(\{X_t = k\}) \quad = \quad \frac{e^{-\lambda_1 tk}(\lambda_1 t)^k}{(\lambda_1 t)!}$$

now suppose that the random variable $\frac{d\mathbb{Q}}{d\mathbb{P}}$ exists and is defined by $\frac{d\mathbb{Q}}{d\mathbb{P}} := \sum_{i=0}^{\infty} z_i 1_{\{X_t=k\}}$ for some given $z_i \in \mathbb{R}$. Then we obtain probabilities for the process under $\mathbb{Q}$ are given by the

Radon-Nikodym theorem as

$$\begin{aligned}\mathbb{Q}(\{X_t = k\}) &= \int_{\{X_t=k\}} \frac{d\mathbb{Q}}{d\mathbb{P}}(\omega) d\mathbb{P}(\omega) \\ &= \int_{\{X_t=k\}} \sum_{i=0}^{\infty} z_i 1_{\{X_t=k\}}(\omega) d\mathbb{P}(\omega) \\ &= \sum_{i=0}^{\infty} z_i \int_{\{X_t=k\}} 1_{\{X_t=k\}}(\omega) d\mathbb{P}(\omega) \\ &= z_k \mathbb{P}(\{X_t = k\})\end{aligned}$$

where the change of limit (the infinite sum) and integral follows from monotone convergence. Now if for $k = 0, 1, \ldots$ we set

$$z_k := e^{(\lambda_2-\lambda_1)t} \left(\frac{\lambda_2}{\lambda_1}\right)^k$$

then the probabilities under $\mathbb{Q}$ are Poisson with parameter λ_2. We have thus found a procedure to move from one probability measure to an equivalent probability measure under which the Poisson process changes its intensity parameter. Sato [1999] gives conditions (Theorems 33.1, 33.2) for general Lévy processes under which this measure change is possible.

2.7.1 The Esscher transform

Let L_t be a Lévy process with characteristic triplet (σ^2, ν, γ) and θ a real number. Assume that ν is such that $\int_{|x|\geq 1} e^{\theta x}\nu(dx) < \infty$. Then an equivalent measure change is defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{\theta L_t}}{\mathbb{E}[e^{\theta L_t}]} = \exp(\theta L_t - \psi(\theta)t). \tag{2.26}$$

where ψ is the characteristic exponent of L_t. Under the new measure $\mathbb{Q}$, L_t is a Lévy process with characteristic triplet $(0, \nu', \gamma')$, with $\nu'(x) = e^{\theta x}\nu(x)$ and $\gamma' = \gamma + \int_{-1}^{1} x(e^{\theta x} - 1)\nu(x)dx$.

Practical implementation in the COS method

In the COS method we use the characteristic function of the density to price options. Therefore if we could find a way to alter the characteristic function to achieve the change of measure, that would be useful to use the Esscher transform in practice. Starting with a probability density $f(x)$ under $\mathbb{P}$, the Esscher transform defines a new probability density $f_\theta(x)$ under $\mathbb{Q}$ as

$$f_\theta(x) = \frac{e^{\theta x} f(x)}{\int e^{\theta x} f(x)dx}. \tag{2.27}$$

It is easily checked that if $\int e^{\theta x} f(x)dx < \infty$, then f_θ is a probability density: it is positive and has integral one. Transform the above to the characteristic function and use the translation theorem:

$$\phi_\theta(u) = \frac{\int e^{iux} e^{\theta x} f(x)dx}{\int e^{\theta x} f(x)dx} = \frac{\int e^{i(u-i\theta)x} f(x)dx}{\int e^{\theta x} f(x)dx} = \frac{\phi(u - i\theta)}{\int e^{\theta x} f(x)dx}. \tag{2.28}$$

where $\phi(u)$ is the characteristic function corresponding to $f(x)$. Of course the correction term $\int e^{\theta x} f(x)dx$ also needs to be expressed in terms of the characteristic function. Applying the

translation theorem again, we can replace $e^{\theta x}f(x) = \frac{1}{2\pi}\int e^{-iux}\phi(u-i\theta)du$ to obtain the correction integral in terms of the characteristic function

$$\int e^{\theta x}f(x)dx \quad = \quad \frac{1}{2\pi}\iint e^{-iux}\phi(u-i\theta)dudx. \tag{2.29}$$

However we are not completely satistied as this double integral will be hard to evaluate numerically. In order to simplify this, let us assume that we have a pure jump Lévy process. Therefore, the characteristic function is

$$\phi(u) \quad = \quad \exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx\right) \tag{2.30}$$

and substituting this into equation (2.29) yields

$$\int e^{\theta x}f(x)dx \quad = \quad \frac{1}{2\pi}\iint e^{-iux}\exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{i(u-i\theta)x}-1)\nu(x)dx\right)dudx. \tag{2.31}$$

Now expand the integral over the Lévy density

$$\int e^{\theta x}f(x)dx \quad = \quad \frac{1}{2\pi}\iint e^{-iux}\exp\left(t\left(\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx+\int_{\mathbb{R}\setminus\{0\}}(e^{\theta x}-1)\nu(x)dx\right)\right)dudx \tag{2.32}$$

and realize that we have just split the characteristic function from a constant term C_θ defined by

$$C_\theta := \exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{\theta x}-1)\nu(x)dx\right). \tag{2.33}$$

We can extract this now from the double integral and realize that the double integral term vanishes:

$$\int e^{\theta x}f(x)dx \quad = \quad \frac{C_\theta}{2\pi}\iint e^{-iux}\phi(u)dudx = C_\theta. \tag{2.34}$$

This yields the result for the characteristic function under the transformed measure $\mathbb{Q}$:

$$\phi_\theta(u) = \frac{\phi(u-i\theta)}{\exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{\theta x}-1)\nu(x)dx\right)} \tag{2.35}$$

The integral that is evaluated in the denominator can be evaluated efficiently (depending on the Lévy density ν) using numerical integration methods (such as a Gaussian quadrature or a Double Exponential integration). It is important to realize that there are some limitations with respect to the choice of θ. In terms of the Lévy density, this limitation corresponds to $\int_{\mathbb{R}\setminus\{0\}}(e^{\theta x}-1)\nu(x)dx < \infty$ as seen in equation (2.35). As an example, when we consider the CGMY model, the G and M parameters limit the choice of θ.

2.7.2 The Mean Correcting Martingale Measure

Schoutens [2003] uses a very practical approach to a risk neutral measure. His approach is to change the drift of the process to the risk free interest rate:

$$S_t' \quad = \quad S_t + mt \tag{2.36}$$

where he sets $m := r - \mathbb{E}_{\mathbb{P}}[S_t]$. This guarantees the martingale property of the discounted stock price under $\mathbb{Q}$. However, it does not provide for the equivalence $\mathbb{Q} \sim \mathbb{P}$. Hence we are not guaranteed an arbitrage free pricing rule. A recent article (Yao et al. [2011]) however shows that under certain conditions on the Lévy measure, the risk neutral measure is indeed equivalent to $\mathbb{P}$, and the prices are indeed free of arbitrage.

A second drawback is that we have no guarantee that the process if defined under $\mathbb{Q}$ (what we do in practice), is still an exponential Lévy process under $\mathbb{P}$. However it is so easy to implement in practice that we will investigate the properties of this measure change.

Practical implementation in the COS method

In order to use the COS-method, we alter the characteristic function $\phi(u)$ of the probability density $f(x)$. First we define $f^*(x) := f(x - m_t)$, with $m_t := \mathbb{E}_{\mathbb{P}}[S_t]$. Then the translation theorem gives us the characteristic function under $\mathbb{Q}$:

$$\phi^*(u) \quad = \quad \phi(u)e^{ium_t}. \tag{2.37}$$

2.7.3 The non uniqueness of the risk neutral measure

In the Black Scholes model, we were used to having a unique equivalent martingale measure (as described in Girsanov's theorem) that transforms one geometric Brownian motion into another (with a different drift parameter). The uniqueness of this risk neutral measure $\mathbb{Q}$ guarantees a unique arbitrage free pricing rule, and thus unique option prices.

In the Lévy framework the risk neutral measure is not unique: in fact there exist infinitely many of them. The cause of this is the existence of jumps in the Lévy framework. In the Black-Scholes framework, the option pricing model is based on the assumption that we can construct a risk free portfolio in which we perfectly replicate the value of an option with a combination of buying/selling the stock and lending/borrowing money at the risk free rate. As there is no uncertainty about this hedging procedure, prices of options are uniquely determined. In the Lévy model we cannot perfectly replicate the option price as we face the risk of a jump occurring in the stock price and not being able to hedge that risk. This means that there can be multiple ways to price options in an arbitrage free way.

From a practical perspective, the choice of the measure might be important from the perspective of parameter stability. It is possible that under one measure transformation the evolution of market prices of risk is smooth with respect to the parameters of the chosen model, while under another measure transformation this is not true at all. We will investigate this in the calibration chapter.

2.8 Hedging in incomplete markets

In incomplete markets, such as markets that contains assets that are modeled by (exponential) Lévy processes, the hedging problem is different than as it is set up in the Black Scholes model. The existence of jumps prevents perfect replication to be possible. We shall first derive the hedging strategy in the Black Scholes model, and then with the guidance of Tankov [2007] we shall explain how the process works in a Lévy framework.

Delta hedging in the Black Scholes model

In the Black Scholes model, the stock price S_t is modeled by an stochastic differential equation

$$dS_t = \mu S dt + \sigma S dW_t \tag{2.38}$$

which is in fact an Itô process. If we define the price of an option $V(t, S_t)$, and assume that it is twice differentiable, Itô's formula gives the dynamics of V

$$dV_t = \left(\mu S \frac{\partial V}{\partial S} + \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt + \sigma S dW_t. \tag{2.39}$$

Now suppose we set up a portfolio Π_t that is short one option and long ϕ_t shares. The dynamics of Π_t are

$$d\Pi_t = -dV + \phi_t dS \tag{2.40}$$

$$= \left(\mu\phi_t S - \mu S \frac{\partial V}{\partial S} - \frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}\right) dt + \sigma S \left(\phi_t - \frac{\partial V}{\partial S}\right) dW_t. \tag{2.41}$$

If we now set $\phi_t := \frac{\partial V}{\partial S}$, the stochastic term disappears in the equation, and the equation becomes deterministic. But as there is no risk associated with it, the portfolio should grow with the risk free interest rate (as otherwise there would be an arbitrage opportunity), and we obtain the equation

$$-\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = r\left(-V + \frac{\partial V}{\partial S} S\right) \tag{2.42}$$

which is the Black Scholes PDE from which we can obtain the Black-Scholes formula for option prices if we take into account the terminal condition $V(T, S_T) = (S_T - K)^+$ in the case of a European call option.

In the meantime, we have found a trading strategy $\phi_t = \frac{\partial V}{\partial S}$ that allows us to set up a portfolio that hedges the risk of the option by trading in the stock. As the portfolio carries no risk, we have created a perfect hedge for the option.

Mean-variance hedging

In a market with Lévy dynamics, the existence of jumps has important implications. Suppose a jump in the stock price ΔS_t occurs at time t, and that we are holding a portfolio that is short one option $C(t, S_t)$ and long ϕ_t stock. The change in the value of the portfolio is

$$\Pi_t - \Pi_{t-} = -\left(C(t, S_t + \Delta S_t) - C(t, S_t)\right) + \phi_t \Delta S_t. \tag{2.43}$$

In order for the hedge strategy ϕ_t to work, the change should be zero, or

$$\phi_t = \frac{C(t, S_t + \Delta S_t) - C(t, S_t)}{\Delta S_t}. \tag{2.44}$$

However the delta hedging strategy $\phi_t = \frac{\partial V}{\partial S}$ does not suffice, as

$$\frac{\partial V}{\partial S} \neq \frac{C(t, S_t + \Delta S_t) - C(t, S_t)}{\Delta S_t}. \tag{2.45}$$

In fact, we cannot find a strategy ϕ_t, trading in the stock only, that completely eliminates the risk associated with the option. The hedging problem then changes shape: we could try to find a

strategy that minimizes the expected variance between our hedging portfolio V_T and the payoff of the option H_T at time T: $\mathbb{E}[(V_T - H_T)^2]$. Ideally we would like to compute this expectation under the real world probabilities $\mathbb{P}$, but since we do not know this in general, we resort to computing the expectation under the risk neutral measure $\mathbb{Q}$. Cont et al. [2007] derives the following result.

$$\mathbb{E}[(V_T - H_T)^2] = (e^{rT}V_0 - \mathbb{E}[H_T])^2 + \mathbb{E}\left[\int_0^T (e^{rT}S_t)^2\sigma^2\left\{\phi_t - \frac{\partial V}{\partial S}\right\}^2 dt\right] \quad (2.46)$$

$$+\mathbb{E}\left[\int_0^T \int_{\mathbb{R}} e^{2r(T-t)}\{C(t, S_t(1+z)) - C(t, S_t) - S_t\phi_t z\}^2 d\nu(x)\right] \quad (2.47)$$

where the last integral is a stochastic integral with respect to the jump measure ν. They also show that the strategy ϕ_t that minimizes this expectation is

$$\phi_t = \frac{\sigma^2\frac{\partial V}{\partial S} + \frac{1}{S_t}\int z(C(t, S_t(1+z)) - C(t, S_t))\nu(z)dz}{\sigma^2 + \int z^2\nu(z)dz}. \quad (2.48)$$

We notice that if there are no jumps present $\nu = 0$, then this is equal to the delta hedging strategy. There are two final comments that are of interest if we wish to hedge options in practice.

First, the above derivation assumed that we only wish to hedge with the stock. In practice we can also hedge with other options to reduce the risk associated with our position. In fact it is the gamma risk (the risk of a changing delta) that we are concerned about.

There is no guarantee that minimizing the variance under $\mathbb{Q}$ also minimizes variance under $\mathbb{P}$. In practice, we might find a suitable measure change $\frac{d\mathbb{P}}{d\mathbb{Q}}$ by analyzing historical data to guess an estimate for $\mathbb{P}$ from the market-fitted risk neutral probabilities $\mathbb{Q}$. Then we can try to minimize the variance under $\mathbb{P}$.

Chapter 3

Option pricing: the COS-method

In Fang and Oosterlee [2008] a method has been developed to efficiently price derivatives using the Fourier cosine transform of the density of the underlying stock price. In a Lévy context this is useful since we do not in general have a density function of the process. We will introduce the method using the notation from Fang and Oosterlee [2008], discuss the three types of errors that are introduced in this approximation method, and discuss the convergence properties of some known processes such as CGMY.

3.1 The method

First of all, for a density function $f : \mathbb{R} \to [0, \infty)$ supported on $[a, b]$, its Fourier cosine expansion is defined as

$$f(x) \quad = \quad {\sum_{k=0}^{\infty}}' A_k \cos\left(k\pi \frac{x-a}{b-a}\right)$$

where the apostrophe means that the first term is counted half. The Fourier cosine coefficients are defined as

$$A_k \quad = \quad \frac{2}{b-a} \int_a^b f(x) \cos\left(k\pi \frac{x-a}{b-a}\right) dx.$$

The risk-neutral pricing formula for an option with payoff $v(S_T, T)$ at time T is

$$v(t, S_t) \quad = \quad e^{-r(T-t)} \int_{\mathbb{R}} v(T, y) f(y) dy. \tag{3.1}$$

Truncating the integration interval and plugging in the cosine approximation for the density (3.4) yields the approximation v_1:

$$\begin{aligned} v_1(t, S_t) &= e^{-r(T-t)} \int_a^b v(T, y) {\sum_{k=0}^{\infty}}' A_k \cos\left(k\pi \frac{y-a}{b-a}\right) dy \\ &= e^{-r(T-t)} {\sum_{k=0}^{\infty}}' A_k \int_a^b v(T, y) \cos\left(k\pi \frac{y-a}{b-a}\right) dy. \end{aligned}$$

Interestingly, the above integral is just $\frac{b-a}{2}$ times the coefficients of the Fourier cosine expansion of the payoff function. Denote V_k as these coefficients:

$$V_k \quad := \quad \frac{2}{b-a} \int_a^b v(T, y) \cos\left(k\pi \frac{y-a}{b-a}\right) dy.$$

Let us assume for now that we can solve this integral (this will depend on the type of the option). The option price then becomes:

$$v_1(t,S_t) = \frac{1}{2}(b-a)e^{-r(T-t)}{\sum_{k=0}^{\infty}}' A_k V_k. \tag{3.2}$$

We can approximate this by truncating the infinite sum:

$$v_2(t,S_t) = \frac{1}{2}(b-a)e^{-r(T-t)}{\sum_{k=0}^{N}}' A_k V_k. \tag{3.3}$$

We now work a bit on the coefficients A_k to obtain them in terms of the characteristic function of $f(x)$. Notice that the coefficients are real since f is a real-valued function:

$$\begin{aligned}
A_k &= \Re\{A_k\} = \Re\left\{\frac{2}{b-a}\int_a^b f(x)\cos\left(k\pi\frac{x-a}{b-a}\right)dx\right\}\\
&= \frac{2}{b-a}\int_a^b f(x)\Re\left\{\cos\left(k\pi\frac{x-a}{b-a}\right)\right\}dx\\
&= \frac{2}{b-a}\int_a^b f(x)\Re\left\{\exp\left(ik\pi\frac{x-a}{b-a}\right)\right\}dx\\
&= \frac{2}{b-a}\Re\left\{\int_a^b f(x)\exp\left(ik\pi\frac{x}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)dx\right\}\\
&= \frac{2}{b-a}\Re\left\{\int_a^b f(x)\exp\left(ik\pi\frac{x}{b-a}\right)dx\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}\\
&\approx \frac{2}{b-a}\Re\left\{\int_{-\infty}^{\infty} f(x)\exp\left(ik\pi\frac{x}{b-a}\right)dx\exp\left(-ik\pi\frac{a}{b-a}\right)\right\}\\
&= \frac{2}{b-a}\Re\left\{\phi\left(\frac{k\pi}{b-a}\right)\exp\left(-ik\pi\frac{a}{b-a}\right)\right\} := F_k.
\end{aligned}$$

We can now replace A_k by F_k in the approximation v_2 to obtain an expression that depends on the characteristic function ϕ (which is known) to obtain the final approximation:

$$v_3(t,S_t) := e^{-r(T-t)}{\sum_{k=0}^{N}}' F_k V_k. \tag{3.4}$$

The characteristic function ϕ can be written as $\phi(u) = e^{iux}\hat{\phi}(u)$, and the coefficients F_k then become

$$F_k = \frac{2}{b-a}\Re\left\{\hat{\phi}\left(\frac{k\pi}{b-a}\right)\exp\left(ik\pi\frac{x-a}{b-a}\right)\right\}. \tag{3.5}$$

The advantage of this notation is that when we wish to price multiple strikes, we only need to calculate the characteristic function once. In this price approximation, three errors were introduced, caused by: the truncated integration interval $[a,b]$, the truncated Fourier cosine series by N, and the replacement of the coefficients A_k by F_k. We need to be able to control these errors to obtain correct option prices.

3.2 Error analysis

In our approximation we have taken three steps. Let us break down the error of the approximation:

$$\begin{aligned}\epsilon &= |v_3(t,S_t) - v(t,S_t)| \\ &\leq |v_3(t,S_t) - v_2(t,S_t)| + |v_2(t,S_t) - v_1(t,S_t)| + |v_1(t,S_t) - v(t,S_t)| \\ &:= \epsilon_3 + \epsilon_2 + \epsilon_1.\end{aligned}$$

Let us discuss each error separately. The first error relates to the truncated integration interval $[a,b]$:

$$\begin{aligned}\epsilon_1 &= |v_1(t,S_t) - v(t,S_t)| = e^{-r(T-t)} \left| \int_a^b v(T,y)f(y)dy - \int_{\mathbb{R}} v(T,y)f(y)dy \right| \\ &= e^{-r(T-t)} \left| \int_{\mathbb{R}\setminus[a,b]} v(T,y)f(y)dy \right|.\end{aligned}$$

The second error is:

$$\epsilon_2 = |v_2(t,S_t) - v_1(t,S_t)| = \frac{1}{2}(b-a)e^{-r(T-t)} \left| \sum_{k=N+1}^{\infty} F_k V_k \right|.$$

The third error relates to the substitution of the coefficients A_k by F_k:

$$\epsilon_3 = |v_3(t,S_t) - v_2(t,S_t)| = e^{-r(T-t)} \left| \sum_{k=0}^{N\,\prime} \left(F_k - \frac{1}{b-a} A_k \right) V_k \right|$$

where

$$\begin{aligned}F_k - \frac{1}{b-a}A_k &= \Re\left\{ \int_{\mathbb{R}\setminus[a,b]} f(x) \exp\left(ik\pi \frac{x}{b-a} \right) dx \exp\left(-ik\pi \frac{a}{b-a} \right) \right\} \\ &= \Re\left\{ \int_{\mathbb{R}\setminus[a,b]} f(x) \exp\left(ik\pi \frac{x-a}{b-a} \right) dx \right\} \\ &= \int_{\mathbb{R}\setminus[a,b]} f(x) \cos\left(k\pi \frac{x-a}{b-a} \right) dx.\end{aligned}$$

Hence

$$\epsilon_3 = |v_3(t,S_t) - v_2(t,S_t)| = e^{-r(T-t)} \left| \sum_{k=0}^{N\,\prime} \left(\int_{\mathbb{R}\setminus[a,b]} f(x) \cos\left(k\pi \frac{x-a}{b-a} \right) dx \right) V_k \right|.$$

The errors can be summarized as follows.

Name	Functional form	Description
ϵ_1	$e^{-r(T-t)} \left\| \int_{\mathbb{R}\setminus[a,b]} v(T,y)f(y)dy \right\|$	integration range error
ϵ_2	$e^{-r(T-t)} \frac{1}{2}(b-a) \left\| \sum_{k=N+1}^{\infty} F_k V_k \right\|$	series truncation error
ϵ_3	$e^{-r(T-t)} \left\| \sum_{k=0}^{N} \left(\int_{\mathbb{R}\setminus[a,b]} f(x) \cos\left(k\pi \frac{x-a}{b-a} \right) dx \right) V_k \right\|$	characteristic replacement error

Increasing the length of the interval $L = b - a$ will **decrease** ϵ_1, **increase** ϵ_2 and **decrease** ϵ_3.

Increasing the cutoff value N has **no effect** on ϵ_1, **decreases** ϵ_2, and has an **undetermined (small) effect** on ϵ_3. Essentially, we should find a balance between the length of the interval L and the truncation value N.

3.2.1 Increasing the length of the integration interval $[a, b]$

Suppose that we have estimated a price using an integration interval $[a^{(1)}, b^{(1)}]$ of length $L = b^{(1)} - a^{(1)}$. and a series truncation value $N^{(1)}$. This yields an error $\epsilon^{(1)} = \epsilon_1^{(1)} + \epsilon_2^{(1)} + \epsilon_3^{(1)}$. Suppose the *accuracy* A is defined as $A = |\epsilon|$. When we increase the integration interval, say to $[a^{(2)}, b^{(2)}]$ of length cL, then what would be the truncation value $N^{(2)}$ needed to reach at least the same accuracy? Since we know that a change in the integration interval only increases the second error ϵ_2, we can deduce that the difference between the new and the old error is bounded by the difference between the second components of the error:

$$|\epsilon_2^{(2)}| - |\epsilon_2^{(1)}| = \left| \frac{1}{2} c L^{(1)} \sum_{k=N^{(2)}+1}^{\infty} F_k V_k \right| - \left| \frac{1}{2} L^{(1)} \sum_{k=N^{(1)}+1}^{\infty} F_k V_k \right| \leq 0.$$

This holds if:

$$\begin{aligned} c \left| \sum_{k=N^{(2)}+1}^{\infty} F_k V_k \right| &\leq \left| \sum_{k=N^{(1)}+1}^{\infty} F_k V_k \right| = \left| \sum_{k=N^{(1)}+1}^{N^{(2)}} F_k V_k + \sum_{k=N^{(2)}+1}^{\infty} F_k V_k \right| \\ &\leq \left| \sum_{k=N^{(1)}+1}^{N^{(2)}} F_k V_k \right| + \left| \sum_{k=N^{(2)}+1}^{\infty} F_k V_k \right|. \end{aligned}$$

We arrive at the easily understood inequality:

$$\underbrace{(c-1) \left| \sum_{k=N^{(2)}+1}^{\infty} F_k V_k \right|}_{\text{error increase due to larger interval}} \leq \underbrace{\left| \sum_{k=N^{(1)}+1}^{N^{(2)}} F_k V_k \right|}_{\text{error decrease due to higher cutoff value}}.$$

This certainly holds if

$$(c-1) \sum_{k=N^{(2)}+1}^{\infty} |F_k||V_k| \leq \sum_{k=N^{(1)}+1}^{N^{(2)}} |F_k||V_k|. \tag{3.6}$$

Inequality (3.6) shows how the error is dependent on the decay of the Fourier cosine coefficients F_k and V_k of the density function and the payoff function respectively.

Since both $|F_k| \leq 1$ and $|V_k| \leq 1$, we have that for any sum over the Fourier coefficients:

$$\sum_k |F_k||V_k| \leq \min \left(\sum_k |F_k|, \sum_k |V_k| \right).$$

Since payoff functions are usually not smooth, they have slowly decaying Fourier coefficients V_k and as such, we usually encounter that the minimum of the above will be the sum of the Fourier coefficients of the density function F_k. In the case of European options modeled by the log-stock price, the payoff functions are actually smoothened by the transformation to the logarithm of the stock price, hence the coefficients decay faster than before.

3.2.2 Increasing the series truncation value N

Increasing the truncation value N from $N^{(1)}$ to $N^{(2)}$ might theoretically cause an increase in the third component of the error:

$$\begin{aligned}
\epsilon_3^{(2)} - \epsilon_3^{(1)} &= \left| \sum_{k=N^{(1)}}^{N^{(2)}} \int_{\mathbb{R}\setminus[a,b]} f(x) \cos\left(k\pi \frac{x-a}{b-a}\right) dx V_k \right| \\
&\leq \sum_{k=N^{(1)}}^{N^{(2)}} \left| \int_{\mathbb{R}\setminus[a,b]} f(x) \cos\left(k\pi \frac{x-a}{b-a}\right) dx \right| |V_k| \\
&\leq \sum_{k=N^{(1)}}^{N^{(2)}} \left| \int_{\mathbb{R}\setminus[a,b]} f(x) dx \right| |V_k| \\
&= \int_{\mathbb{R}\setminus[a,b]} f(x) dx \sum_{k=N^{(1)}}^{N^{(2)}} |V_k| \,.
\end{aligned}$$

This error is small compared to the other errors, since if the interval $[a, b]$ is reasonably large, the integral $\int_{\mathbb{R}\setminus[a,b]} f(x)dx$ will be small as well as the sum of the Fourier coefficients will be small as $N^(1)$ will be large. In fact, Fang and Oosterlee [2008] shows that $|\epsilon_3| < |\epsilon_1| + O\left(\int_{\mathbb{R}\setminus[a,b]} f(x)dx\right)$.

3.2.3 The integration interval error ϵ_1 cannot be bounded for arbitrary Lévy processes and payoff functions

In principle a probabililty density could have all its mass outside any interval $[a, b]$, and the payoff function $v(T, y)$ could grow indefinately outside of $[a, b]$, so we cannot make any statements on the absolute size of the integration interval error ϵ_1. We should assume both a model for the probability density and the payoff to estimate this error. We will try to give bounds for the CGMY process and for European options.

3.2.4 Bounding the integration interval error ϵ_1 for the European call option and a density function with exponential tail decay

The error ϵ_1 introduced by the integration interval $[a, b]$ reads

$$\epsilon_1 = e^{-r(T-t)} \left| \int_{\mathbb{R}\setminus[a,b]} v(T, y) f(y) dy \right| .$$

We will assume that the density has some exponential decay with parameter R in its tail after $x = b$:

$$f(x) \leq f(b) \exp\left(-R(x-b)\right) \quad \text{for some } R > 1.$$

Mind that if we model the distribution of the log-stock price, this assumption corresponds exactly with the distribution S_T having a finite Rth moment. Whether this holds for any $R > 1$ is not given, but should be determined for each model. For the CGMY model it for instance holds if for the M parameter, we have $M < R$. We can deduce that

$$\int_b^\infty f(x)dx \leq \frac{f(b)}{R} \tag{3.7}$$

and

$$\int_b^\infty e^x f(x)dx \leq \frac{f(b)}{R-1}e^{bR}. \tag{3.8}$$

Subtracting (3.7) from (3.8) yields the error bound

$$\int_b^\infty (e^x - 1)f(x)dx \leq f(b)\left(\frac{e^{bR}}{R-1} - \frac{1}{R}\right).$$

This implies that if $b > 0$ is the right side of our integration interval $[a, b]$, we have the error bound for a European call option:

$$\epsilon_1 \quad = \quad e^{-r(T-t)}\left|\int_b^\infty K(e^x - 1)^+ f(x)dx\right| \leq e^{-r(T-t)}f(b)K\left(\frac{e^b}{R-1} - \frac{1}{R}\right).$$

This bound seems very simple, but it has two drawbacks.

- Its crucial assumption is exponential decay in the tail of the density function. This might however be deduced from the knowledge of the Lévy density of the process, since the assumption corresponds to the Rth moment of the exponential Lévy process being finite. For CGMY, this is determined by the M parameter.
- It uses a value $f(b)$ of the density function which we do not know in general. In practice we should estimate this value, which introduces another error.

3.2.5 Bounding the series truncation error ϵ_2

The bound of the series truncation error depends on the decay of the Fourier coefficients of the payoff and the density. As the payoff usually is not smooth, we assume that the coefficients of the denstiy decay faster and use this as a bound. Use equation (3.13) to obtain

$$\begin{aligned}
\epsilon_2 \quad &= \quad \left|\frac{1}{2}(b-a)e^{-r(T-t)}\sum_{k=N+1}^{\infty} F_k V_k\right| \leq \frac{1}{2}(b-a)e^{-r(T-t)}\sum_{k=N+1}^{\infty}|F_k||V_k| \\
&\leq \quad \frac{1}{2}(b-a)e^{-r(T-t)}\sum_{k=N+1}^{\infty}|F_k||V_k| \leq \frac{1}{2}(b-a)e^{-r(T-t)}\sum_{k=N+1}^{\infty}|F_k| \\
&\leq \quad e^{-r(T-t)}\sum_{k=N+1}^{\infty}\left|\phi\left(\frac{k\pi}{b-a}\right)\right|. \qquad (3.9)
\end{aligned}$$

The size of this error will depend on the particular process used. Let us investigate the case of the CGMY process.

Bounding ϵ_2 for the CGMY process

For the CGMY process, we know that the absolute of the characteristic function is exponentially bounded:

$$|\phi(u)| \leq \widetilde{A}\exp\left(-\widetilde{B}u^Y\right).$$

This implies that for the Fourier cosine coefficients:

$$\begin{aligned}
|F_k| \quad &\leq \quad \left|\phi\left(\frac{k\pi}{b-a}\right)\right| = \frac{2}{b-a}\widetilde{A}\exp\left(-\widetilde{B}\left(\frac{k\pi}{b-a}\right)^Y\right) \qquad (3.10) \\
&= \quad \frac{2}{b-a}\widetilde{A}\exp\left(-\widetilde{R}k^Y\right) \qquad (3.11)
\end{aligned}$$

where $\widetilde{R} := \widetilde{B}\left(\frac{\pi}{b-a}\right)^Y$. An infinite sum of exponentials can be bounded by an integral since it is a decreasing function:

$$\sum_{k=N+1}^{\infty} \exp\left(-\widetilde{R}k^Y\right) \leq \exp\left(-\widetilde{R}(N+1)^Y\right) + \int_{N+1}^{\infty} \exp\left(-\widetilde{R}x^Y\right) dx.$$

This integral leads to a special function: the incomplete Gamma function $\Gamma_{a,b}(s)$. Substitute $z := \widetilde{B}x^Y$:

$$\int_{N+1}^{\infty} \exp\left(-\widetilde{R}x^Y\right) dx = \frac{1}{\widetilde{R}^{1/Y}Y} \int_{\widetilde{R}(N+1)^Y}^{\infty} z^{\frac{1}{Y}-1} \exp(-z)\, dz = \frac{1}{\widetilde{R}^{1/Y}Y} \Gamma_{\widetilde{R}(N+1)^Y,\infty}\left(\frac{1}{Y}\right).$$

Hence the bound becomes

$$\begin{aligned}\sum_{k=N+1}^{\infty} |F_k| &\leq \frac{2}{b-a}\widetilde{A} \sum_{k=N+1}^{\infty} \exp\left(-\widetilde{R}k^Y\right) \\ &\leq \frac{2}{b-a}\widetilde{A}\left[\exp\left(-\widetilde{R}(N+1)^Y\right) + \frac{1}{\widetilde{R}^{1/Y}Y}\Gamma_{\widetilde{R}(N+1)^Y,\infty}\left(\frac{1}{Y}\right)\right].\end{aligned}$$

Concluding, the series truncation error is bounded (recalling equation (3.9)) by

$$\epsilon_2 \leq e^{-r(T-t)}\widetilde{A}\left[\exp\left(-\widetilde{R}(N+1)^Y\right) + \frac{1}{\widetilde{R}^{1/Y}Y}\Gamma_{\widetilde{R}(N+1)^Y,\infty}\left(\frac{1}{Y}\right)\right]. \tag{3.12}$$

3.2.6 Bounding ϵ_3

The second error is due to replacing A_k by F_k:

$$\epsilon_3 = e^{-r(T-t)} \sum_{k=0}^{N-1'} V_k \int_{\mathbb{R}\setminus[a,b]} \cos\left(k\pi\frac{y-a}{b-a}\right) f(y)dy.$$

Following Fang and Oosterlee [2008], we can express this in terms of ϵ_1:

$$\begin{aligned}\epsilon_3 &= e^{-r(T-t)} \int_{\mathbb{R}\setminus[a,b]} \left[v(y) - \sum_{k=N}^{\infty} V_k \cos\left(k\pi\frac{y-a}{b-a}\right)\right] f(y)dy \\ &= \epsilon_1 - e^{-r(T-t)} \int_{\mathbb{R}\setminus[a,b]} \left[\sum_{k=N}^{\infty} V_k \cos\left(k\pi\frac{y-a}{b-a}\right)\right] f(y)dy.\end{aligned}$$

Suppose that we could find a bound on the decay of the Fourier coefficients V_k, so that $\sum_{k=N}^{\infty} |V_k| < \infty$. We can then bound the sum inside the brackets:

$$\left|\sum_{k=N}^{\infty} V_k \cos\left(k\pi\frac{y-a}{b-a}\right)\right| \leq \sum_{k=N}^{\infty} |V_k|.$$

This means that for the error

$$|\epsilon_3| \leq |\epsilon_1| + e^{-r(T-t)} \sum_{k=N}^{\infty} |V_k| \int_{\mathbb{R}\setminus[a,b]} f(y)dy.$$

The integral is of course at most one, since it is a density function:

$$|\epsilon_3| \leq |\epsilon_1| + e^{-r(T-t)} \sum_{k=N}^{\infty} |V_k|.$$

3.3 Smoothness and decay of the Fourier series coefficients

The literature on harmonic analysis has long discovered the connection between the decay of Fourier coefficients and the smoothness of the function in the time domain. Suppose a function f is in the Schwartz space $S(\mathbb{R})$. The Fourier coefficients for its derivative (denoted by $\widehat{f'}(k)$) then are

$$\widehat{f'}(k) \quad = \quad ik\widehat{f}(k).$$

Taking the absolute values then yields for $k \neq 0$:

$$|\widehat{f}(k)| = \frac{|\widehat{f'}(k)|}{k} \leq \frac{A}{k} \quad \text{for some } A \in \mathbb{R}.$$

The last inequality follows from $|\widehat{f'}(k)| = |\int_{\mathbb{R}} f(x)e^{iux}dx| \leq \int_{\mathbb{R}} |f(x)|e^{iux}dx \leq \int_{\mathbb{R}} |f(x)|dx = A < \infty$ since $f \in \text{L}^1(\mathbb{R}) \subset S(\mathbb{R})$. This all easily generalizes to higher derivatives. Without proof we note that for the n-th derivative $f^{(n)}$, we have

$$|\widehat{f}(k)| \quad \leq \quad \frac{A}{k^n} \quad \text{for some } A \in \mathbb{R}. \tag{3.13}$$

This might provide us with a method to get a bound on the left hand side of equation (3.6) if we can assume the smoothness implied by the Schwartz space on either the density function or the payoff function.

3.3.1 Exponential decay of the Fourier series coefficients

In the Lévy framework we can encounter characteristic functions that decay exponentially:

$$|\phi(u)| \quad = \quad \exp(-Ku)$$

for some $K > 0$. Note that an exponentially decreasing function will always surpass a polynomially decreasing function at some point. Therefore, for any degree n of a polynomial, there exists an A, such that

$$|\phi(u)| \leq \frac{A}{u^n}.$$

A consequence of this is, that if a characteristic function decays exponentially, all of its derivatives will exist and are in $\text{L}^2(\mathbb{R})$. This means the characteristic function corresponds to a smooth density function.

3.3.2 Connecting the Fourier cosine coefficients to the characteristic function

If we are given any density function $f(x)$ supported on $[a, b]$ we can define its translated version $g(x) := f(x + a)$. Define the Fourier cosine transform of g as

$$g(x) \quad = \quad \sum_{k=0}^{\infty}{}' A_k \cos(k\pi x)$$

with Fourier cosine coefficients

$$A_k \quad = \quad \frac{2}{b-a}\int_a^b g(x)\cos(k\pi x)dx.$$

When $[a, b]$ is wide enough, these (real) coefficients are approximately the Fourier Transform of g evaluated at points $k\pi$:

$$\begin{aligned} A_k &= \frac{2}{b-a}\int_a^b g(x)\cos(k\pi x)dx \approx \frac{2}{b-a}\int_{-\infty}^{\infty} g(x)\cos(k\pi x)dx \\ &= \frac{2}{b-a}\int_{-\infty}^{\infty} g(x)e^{ik\pi x}dx = \frac{2}{b-a}\phi_g(k\pi). \end{aligned}$$

Here ϕ_g is the Fourier Transform of $g(x)$. We can use the translation property of the Fourier transform to obtain the coefficients in terms of the Fourier transform of $f(x)$ now:

$$A_k \approx \frac{2}{b-a}\phi_f(k\pi)e^{iux}.$$

Taking absolutes, we obtain the useful approximation

$$|A_k| \approx \frac{2}{b-a}|\phi_f(k\pi)|.$$

Application: the Poisson process

The characteristic function of the Poisson process with intensity parameter λ is

$$\phi(u) = \exp\lambda\left(e^{iu}-1\right)$$

leading to the absolute value:

$$\begin{aligned} |\phi(u)| &= \exp\lambda\left(\Re\{e^{iu}-1\}\right) \\ &= \exp\lambda\left(\cos(u)-1\right) \end{aligned}$$

This is an oscillating function with values $e^{-2\lambda} \leq |\phi(u)| \leq 1$. Since this characteristic function does not decay at all, the Fourier cosine coefficients do not decay either. It is an example of a Lévy process for which the Fourier cosine series do not converge and we therefore cannot use the COS-method (or any other Fourier-based method for that matter). This is connected to the fact that the Poisson distribution is not continuous.

Application: the Black-Scholes model

The Black-Scholes model (or geometric Brownian motion) has characteristic function

$$\phi(u) = \exp(i\mu u - \frac{1}{2}\sigma^2u^2)$$

and its absolute value reads

$$|\phi(u)| = \exp(-\frac{1}{2}\sigma^2u^2).$$

The characteristic function decays exponentially. This tells us that the density function must be a smooth function.

Application: the CGMY process

The CGMY process has characteristic function

$$\phi(u) = \exp\left(C\Gamma(-Y)\left((M-iu)^Y - M^Y + (G+iu)^Y - G^Y\right)\right).$$

Taking the absolute value yields

$$|\phi(u)| = \exp\left(C\Gamma(-Y)\left(\Re\{(M-iu)^Y\} - M^Y + \Re\{(G+iu)^Y\} - G^Y\right)\right).$$

For $0 < Y < 1$ we have $\Gamma(-Y) < 0$ and for $1 < Y < 2$ we have $\Gamma(-Y) > 0$. At the points $Y = 0, 1, 2$ we have $\Gamma(-Y) = \infty$. Let's see what the above real parts of the exponential look like. For any complex number z, we have the polar form $z = |z|e^{i\theta}$ where $\theta = \arg(z) \in (-\pi, \pi]$. Taking $z := M - iu$ we have

$$M - iu \;=\; |M-iu|e^{i\theta_M(u)} = \sqrt{M^2+u^2}e^{i\theta_M(u)}$$

where $\theta_M(u) = \arg(M-iu)$ can be determined from $\cos(\theta_M(u)) = \frac{M}{\sqrt{M^2+u^2}}$ and $\sin(\theta_M(u)) = -\frac{u}{\sqrt{M^2+u^2}}$. We just note that $\cos(\theta_M(u)) \to 0$ and $\sin(\theta_M(u)) \to -1$ as $u \to \infty$. This implies that $\theta_M \to -\frac{\pi}{2}$ as $u \to \infty$. Now take this to the power Y:

$$(M-iu)^Y \;=\; |M-iu|^Y e^{iY\theta_M(u)} = (M^2+u^2)^{Y/2}e^{iY\theta_M(u)}$$

and take the real part of this:

$$\Re\{(M-iu)^Y\} \;=\; (M^2+u^2)^{Y/2}\cos(Y\theta_M(u)).$$

Since $\theta_M(u) \to -\frac{\pi}{2}$ we can deduce that for the CGMY process

$$\begin{aligned}\lim_{u\to\infty}\cos(Y\theta_M(u)) &> 0 \quad \text{if } 0<Y<1\\ \lim_{u\to\infty}\cos(Y\theta_M(u)) &< 0 \quad \text{if } 1<Y<2.\end{aligned}$$

Now consider $(G+iu)^Y$. Equivalently to the previous, we have

$$\Re\{G+iu\} \;=\; \sqrt{G^2+u^2}\cos\left(\theta_G(u)\right)$$

where $\cos(\theta_G(u)) = \frac{G}{\sqrt{G^2+u^2}}$ and $\sin(\theta_G(u)) = \frac{u}{\sqrt{G^2+u^2}}$. This implies that $\cos(\theta_G(u)) \to 0$ and $\sin(\theta_G(u)) \to 1$, and thus $\theta_G(u) \to \frac{\pi}{2}$ as $u \to \infty$. Taking the power Y of this, we obtain

$$\Re\{(G+iu)^Y\} \;=\; (G^2+u^2)^{Y/2}\cos\left(Y\theta_G(u)\right)$$

and for values of $0 < Y < 2$ we distinguish between:

$$\begin{aligned}\lim_{u\to\infty}\cos(Y\theta_G(u)) &> 0 \quad \text{if } 0<Y<1\\ \lim_{u\to\infty}\cos(Y\theta_G(u)) &< 0 \quad \text{if } 1<Y<2.\end{aligned}$$

We can now return to the absolute of the characteristic function and bound it. Suppose $0 < Y < 1$.

$$\begin{aligned}
|\phi(u)| &= \exp\left(C\Gamma(-Y)\left(\Re\{(M-iu)^Y\} - M^Y + \Re\{(G+iu)^Y\} - G^Y\right)\right) \\
&= A\exp\left(-B\left(\Re\{(M-iu)^Y\} + \Re\{(G+iu)^Y\}\right)\right) \\
&= A\exp\left(-B\left\{(M^2+u^2)^{Y/2}\cos\left(Y\theta_M(u)\right) + (G^2+u^2)^{Y/2}\cos\left(Y\theta_G(u)\right)\right\}\right) \\
&\leq A\exp\left(-B\left\{(M^2+u^2)^{Y/2}\cos\left(Y\frac{\pi}{2}\right) + (G^2+u^2)^{Y/2}\cos\left(-Y\frac{\pi}{2}\right)\right\}\right) \\
&\leq A\exp\left(-B\left\{(M^Y+u^Y)\cos\left(Y\frac{\pi}{2}\right) + (G^Y+u^Y)\cos\left(Y\frac{\pi}{2}\right)\right\}\right) \\
&\leq A\exp\left(-B\left\{(M^Y+G^Y)\cos\left(Y\frac{\pi}{2}\right) + 2u^Y\cos\left(Y\frac{\pi}{2}\right)\right\}\right) \\
&= \widetilde{A}\exp\left(-2Bu^Y\cos\left(Y\frac{\pi}{2}\right)\right) = \widetilde{A}\exp\left(-\widetilde{B}u^Y\right)
\end{aligned}$$

where

$$\begin{aligned}
A &:= \exp\left(-C\Gamma(-Y)\left(M^Y+G^Y\right)\right) \\
B &:= -C\Gamma(-Y) \\
\widetilde{A} &:= A + (M^Y+G^Y)\cos\left(Y\frac{\pi}{2}\right) \\
\widetilde{B} &:= 2B\cos\left(Y\frac{\pi}{2}\right)
\end{aligned}$$

are all positive constants (with respect to u). We have thus given an exponential bound on the decay of the characteristic function (for $0 < Y < 1$ only, however for $1 < Y < 2$ the calculation is the same).

3.4 Error bounds for the European call option under the CGMY process

We shall now summarize the results from the previous section to arrive at an error bound for the European call option under the CGMY process. We assume that the density of the CGMY process has exponential tail decay after b with decay parameter at least G. This is equal to the assumption that the Rth moment of S_T should be finite. Furthermore we assume that the density function is in C^∞. For any $[a,b]$ and N, we have

$$\begin{aligned}
|\epsilon_1| &\leq e^{-r(T-t)}f(b)K\left(\frac{e^{bR}}{G-1} - \frac{1}{G}\right) \\
|\epsilon_2| &\leq e^{-r(T-t)}\widetilde{A}\left[\exp\left(-\widetilde{R}(N+1)^Y\right) + \frac{1}{\widetilde{R}^{1/Y}Y}\Gamma_{\widetilde{R}(N+1)^Y,\infty}\left(\frac{1}{Y}\right)\right] \\
|\epsilon_3| &\leq |\epsilon_1| + \int_{\mathbb{R}\setminus[a,b]} f(x)dx
\end{aligned}$$

using the variables

$$\begin{aligned}
\widetilde{A} &= \exp\left(-C\Gamma(-Y)(M^Y+G^Y)\right) + (M^Y+G^Y)\cos\left(Y\frac{\pi}{2}\right) \\
\widetilde{R} &= -2C\Gamma(-Y)\cos\left(Y\frac{\pi}{2}\right)\left(\frac{\pi}{b-a}\right)^Y.
\end{aligned}$$

Concluding, we obtain the bound

$$|\epsilon| \leq |\epsilon_1| + |\epsilon_2| + |\epsilon_3|.$$

3.5 Convergence of the Fourier cosine expansion and the Fourier transform

The convergence of the Fourier cosine expansion is not guaranteed for any function f, but since we are working on a finite interval $[a, b]$ this usually poses no problems. If f is continuously differentiable, this guarantees uniform convergence of the Fourier cosine expansion. However there are some issues regarding round-off errors when f becomes large.

Cosine expansion of the payoff function: European call option instability

In this thesis we focus on European call options, that are characterized by their payoff function $(S_T - K)^+$ for a call option, and $(K - S_T)^+$ for a put option. Using the change of variables $x = \log\left(\frac{S_T}{K}\right)$ we have $K(e^x - 1)$ for a call option and $K(1 - e^x)^+$ for a put option. We see that the function for the call option is unbounded as $x \to \infty$, but the put option is bounded as $0 \leq x < 1$ for all x:

Figure 3.1: Call option payoff $(e^x - 1)^+$

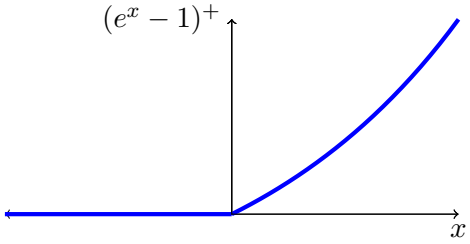
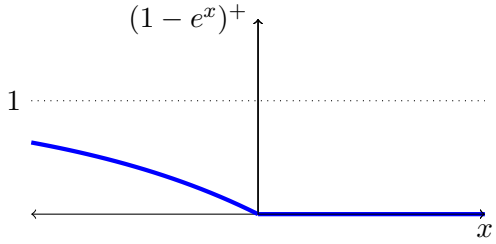


Figure 3.2: Put option payoff $(1 - e^x)^+$



If we now look at the coefficients F_k follow from the call option, we see that as b becomes large, for some fixed k these coefficients can become very large. Using the COS method to price options, this will in theory be compensated by a fast declining part of the probability density. However in a computer with finite precision this will lead to round-off errors. The put option does not suffer from these problems as the function is bounded on $\mathbb{R}$. Therefore we shall price call options using the put-call parity.

We could have avoided the problem at all by having chosen the transformation $x = -\log\left(\frac{S_T}{K}\right)$ instead of $x = \log\left(\frac{S_T}{K}\right)$. Then the call option would be bounded, and we would have moved the problem to the put options. We would have to adjust the characteristic function of the density as well then.

Fourier transform of the density

In the COS method we use the characteristic function of the density. Since for any Lévy process, the characteristic function exists and follows from the Lévy-Khintchine theorem, we are not concerned with its existence or convergence. However, for our practical use it is important that the characteristic function decays fast, as a slow convergence might lead to a large truncation value N to be necessary. As we will see later, the decay of the Fourier transform is connected to the smoothness of the associated probability density. We will see that there are Lévy processes with associated characteristic functions that do not decay at all (such as the Poisson process), and therefore we cannot use them in the COS-method.

3.6 The integration interval: a numerical approximation of the moments and cumulants

Obtaining an integration interval is essential for the execution of the COS-method. Especially for short maturities, the density of the log-stock price can becomes really peaked and we should position it accurately. The method of cumulants as proposed in Fang and Oosterlee [2008] states that we can take an interval in terms of the *cumulants* c_i of the distribution as

$$[a,b] := \left[c_1 - L\sqrt{c_2 + \sqrt{c-4}}, c_1 + L\sqrt{c_2 + \sqrt{c_4}}\right]$$

with L a parameter that is chosen such that the interval is wide enough. Fang and Oosterlee [2008] provides an indication that $L = 10$ should do for most cases. The *cumulant generating function* of a random variable X is given by

$$f(t) \quad := \quad \log\left(\mathbb{E}[e^{tX}]\right)$$

And the k-th cumulant is then defined as the n-th derivative evaluated at zero:

$$c_n \quad := \quad f^{(n)}(0).$$

We can easily see that the for a Lévy process L_t with characteristic function $\phi(u) = \mathbb{E}[e^{tL_1}] = \exp(\psi(iu))$, the cumulant generating function is

$$f(t) = \log\left(\mathbb{E}[e^{tX}]\right) = \log\left(\phi(-it)\right) = \psi(t)$$

and thus equal to what we earlier referred to as the characteristic exponent of the Lévy process. For some well-known processes (such as Variance Gamma or CGMY) the cumulants are known analytically, so that calculation of this interval is easily done. However, in our research we will encounter processes of which we not necessarily have an analytical form for the characteristic function, and as such also not for the cumulants. Therefore we need to find an approximation for the cumulants. Now we are busy with that we might as well obtain the moments of the process first. The cumulants follow from the relationship

$$c_n = m'_n - \sum_{m=1}^{n-1} \binom{n-1}{m-1} c_m m_{n-m}$$

where the m_i are raw moments (not centered). The moments can be calculated from the characteristic function by

$$m_n \quad = \quad \frac{1}{i^n}\phi^{(n)}(0).$$

Now an approach could be that we take a finite difference to approximate the derivative. We would like to use as least evaluations of the characteristic function as possible. Then for the first moment we could try $m_1 \approx \frac{1}{i}\frac{\phi(h)-\phi(0)}{h}$ with some small $h \in \mathbb{R}$. However for complex functions this suffers from round-off errors due to subtractive cancellation errors. A method has been provided in Martins et al. [2003] that gives a similar approach but suited for complex functions

called the *complex step approximation*. The first two moments are given by

$$\begin{aligned}
m_1 &\approx \frac{1-\Re(\phi(ih_1))}{h_1} \\
m_2 &\approx 2\frac{1-\Re(\phi(h_2))}{h_2^2} \\
m_3 &\approx -\frac{\Im\left(\phi(-3h_3)-8\phi(-2h_3)+13\phi(-h_3)-13\phi(h_3)+8\phi(2h_3)-\phi(3h_3)\right)}{8h_3^3} \\
m_4 &\approx \frac{\Re\left(-\phi(-3h_4)+12\phi(-2h_4)-39\phi(-h_4)+56-39\phi(h_4)+12\phi(2h_4)-\phi(3h_4)\right)}{6h_4^4}
\end{aligned}$$

where we take $h_1 := 10^{-8}$, $h_2 := 10^{-4}$, $h_3 := 10^{-3}$, $h_4 := 10^{-2}$.

3.7 The Greeks

Obtaining the Greeks (the sensitivities of the option prices with respect to the model parameters) is useful for hedging purposes (even though delta hedging does not work in incomplete markets - but we will get to that later). In the COS method we work with the log-stock price $x = \log\left(\frac{S_T}{K}\right)$. The delta and gamma can be derived as

$$\Delta = \frac{dv}{dS} = \frac{\partial v}{\partial x}\frac{\partial x}{\partial S} = \frac{1}{S}\frac{\partial v}{\partial x} \tag{3.14}$$

$$\Gamma = \frac{d^2v}{dS^2} = \frac{1}{S^2}\left(\frac{\partial^2 v}{\partial x^2} - \frac{\partial v}{\partial x}\right). \tag{3.15}$$

The COS-approximation for the option,

$$v(x) = e^{-r(T-t)}{\sum_{k=0}}'^{N-1}\Re\left\{\phi(\frac{k\pi}{b-a})\exp(ik\pi\frac{x-a}{b-a})\right\}V_k \tag{3.16}$$

can be differentiated with respect to x to obtain the Greeks as

$$\Delta = \frac{1}{S}\frac{\partial v}{\partial x} = e^{-r(T-t)}{\sum_{k=0}}'^{N-1}\Re\left\{\phi(\frac{k\pi}{b-a})\exp(ik\pi\frac{x-a}{b-a})\frac{ik\pi}{b-a}\frac{1}{S}\right\}V_k, \tag{3.17}$$

$$\Gamma = \frac{1}{S}\frac{\partial v}{\partial x} = e^{-r(T-t)}{\sum_{k=0}}'^{N-1}\Re\left\{\phi(\frac{k\pi}{b-a})\exp(ik\pi\frac{x-a}{b-a})\left[\left(\frac{ik\pi}{b-a}\right)^2 - \frac{ik\pi}{b-a}\right]\frac{1}{S^2}\right\}V_k \tag{3.18}$$

Let us give a numerical example of these sensitivities. Consider a CGMY model fitted to the Jan-2012 DAX call options on 04-Jan-2012 at 11:00. Its parameters read $C = 0.06, G = 3.40, M = 32.1, Y = 1.40$, and the option price, the deltas and the gammas are displayed as

Figure 3.3: Call option price

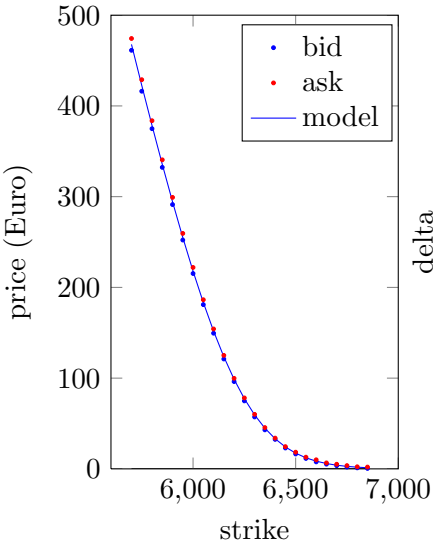


Figure 3.4: Call option delta

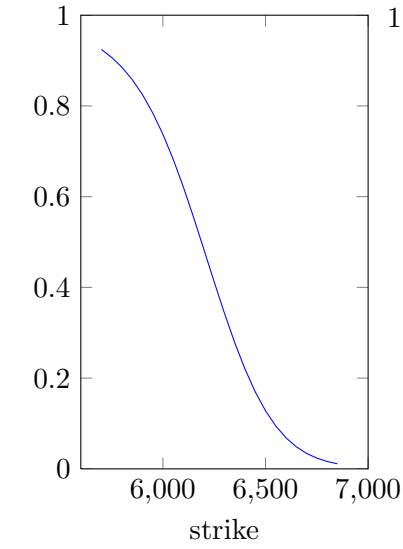
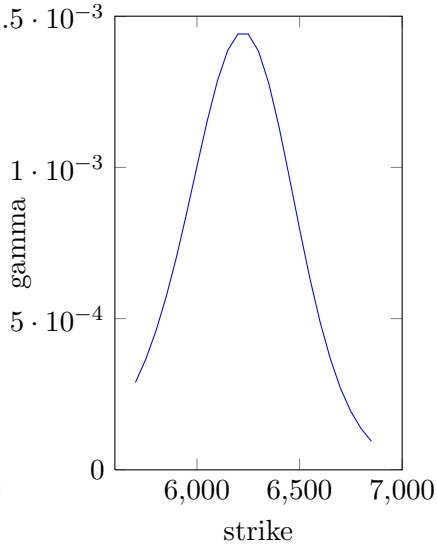


Figure 3.5: Call option gamma



For Lévy models, we do not in general have a vega (the sensivity of the option price with respect to the volatility parameter σ). We can however define some new notion of sensitivity with respect to moments.

Sensitivities with respect to moments

In the Black-Scholes model, we can define the sensitivity with respect to the volatility parameter σ, referred to as the *vega*

$$\text{vega} \quad = \quad \frac{\partial C}{\partial \sigma}$$

which represents how the option price C will change relatively to the change in volatility σ. In Lévy-based models we have not defined a volatility parameter, hence we cannot define this quantity. However, as seen above we can derive the moments of the process at fixed times. Therefore, we could potentially define sensitivities of the stock price with respect to the moments m_n (they could be raw moments, central moments or cumulants): $\frac{\partial C}{\partial m_n}$. Since C is not usually a function of moments, yet of other parameters $p_1, \dots p_m$, we could find its total derivative with respect to m_n as

$$\frac{dC}{dm_n} \quad = \quad \frac{\partial C}{\partial p_1}\frac{\partial p_1}{\partial m_n} + \dots + \frac{\partial C}{\partial p_1}\frac{\partial p_m}{\partial m_n} \tag{3.19}$$

The sensitivities of the option price with respect to the parameters can be easily determined (within the COS method for instance), however the sensitivities of the parameters with respect to the moments are not easily found. Let us make a fundamental assumption. In general we can say that the moments are a function of the parameters:

$$(m_1, \dots, m_n) = f(p_1, \dots, p_m) \tag{3.20}$$

for some known function f. We will assume that the parameters uniquely determine the moments and that an inverse f^{-1} exists such that

$$(p_1, \dots, p_m) = f^{-1}(m_1, \dots, m_n) \tag{3.21}$$

We will assume that

$$\frac{\partial p_m}{\partial m_n} = \left(\frac{\partial m_n}{\partial p_m}\right)^{-1}. \tag{3.22}$$

This implies that the *moment sensitity* $\frac{\partial C}{\partial m_n}$ has become

$$\frac{dC}{dm_n} = \frac{\partial C}{\partial p_1}\left(\frac{\partial m_n}{\partial p_1}\right)^{-1} + \ldots + \frac{\partial C}{\partial p_1}\left(\frac{\partial m_n}{\partial p_m}\right)^{-1}. \tag{3.23}$$

The above procedure could provide us with a sensititivity similar to the vega. As stated in the introduction, the Greeks are important to option traders as they provide a quick insight into the risk profile of an option.

3.8 Conclusion

In this chapter we have introduced the COS-method as developed in Fang and Oosterlee [2008]. While the original article focusses on the convergence asymptotics of the method, we have investigated some properties of the absolute size of the error involved assuming particular models. We have presented a simple way to obtain the cumulants for a process of which we know the characteristic function but have no analytical version of the moments and/or the cumulants.

There are some alternatives to using the COS-method to calculate option prices from the characteristic function. Carr and Madan [1999] uses the Fast Fourier Transform to integrate $\int v(y)f(y)dy$ that is first transformed using the Parseval-Plancherel formula. Fang and Oosterlee [2008] however shows that the COS-method is faster in implementation. For our use the speed is quite important as the calibration to market data requires many evaluations of prices under different parameter settings.

Chapter 4

Calibration procedure

4.1 Input data and the minimization problem

We are given a data set that consists of N bid prices $(b_i)_{i=1}^N$ and ask prices $(a_i)_{i=1}^N$. Each observation $i \in \{1, \ldots, N\}$ has several properties, such as whether it is a call or a put option, a strike price and a maturity. Furthermore we assume a parametric model f with parameter vector $\Theta = (\theta_1, \ldots, \theta_p)^T$ that gives option prices $p_i = f(\Theta, i)$ for each observation i. We wish to determine a parameter vector for which the prices p_i are 'close' to the bid and ask prices. To define this notion we introduce a function $l(b_i, a_i, p_i) \geq 0$ called a *loss function* to define how well a particular observation fits the data. We discuss a few ideas to construct such a function and discuss their properties.

4.2 Filtering of raw observations

Our dataset consists of DAX index options and futures data proprietary to Optiver. We have filtered it, as the dataset contains less informative observations such as containing only a bid price and not an ask price, or even a bid price that is larger than an ask price as a consequence of faulty data. Therefore we have created some rules that filter out any obvious problems.

Nonzero and consistent bid and ask prices

We require for an observation that both the bid and the ask price are nonzero, and that the ask price is strictly larger than the bid price.

Spread constraints

It sometimes happens that market participants are not willing to take on certain risks; for instance sell a far out of the money put option that has a very low theoretical price but imposes a significant loss in the case of an extreme event. It could happen that an ask price for such an option is quoted at a very high level. We regard these as not containing much information for our calibration purposes, and wish to disregard these observations. The rules we have set up are quite arbitrary, but they 'do the job' for us. We require the spread (ask price minus the bid price) to adhere to

1. spread $\leq 3\sqrt{\text{bid}}$
2. spread $\leq 3\sqrt{\text{ask}}$
3. spread $\leq 9\text{bid}^2$

4. spread $\leq 9\text{ask}^2$.

The inclusion of both a square root of the price and a squared price comes their different effect on prices below 1 Euro and above that level.

4.3 Loss functions

Selecting the right loss function is a very difficult problem. Cont and Tankov [2004] discusses this (usually ill posed) problem. In this thesis we will investigate a few different approaches and discuss their properties.

Squared distance to mid

An easy choice is to consider the distance of the price p_i to the *mid price* $m_i := \frac{a_i+b_i}{2}$ and square it:

$$l(b_i, a_i, p_i) := (p_i - m_i)^2 \tag{4.1}$$

The *average loss* is then defined as

$$L := \frac{1}{N}\sum_{i=1}^{N} l(b_i, a_i, p_i). \tag{4.2}$$

We introduce the notion of an *optimum* as a parameter $\hat{\Theta}$ such that

$$L(\hat{\Theta}) = 0 \tag{4.3}$$

A nice property of this loss function is that it defines a unique optimum if it exists. We can deduce this result easily by assuming that an optimum $\hat{\Theta}$ exists. Then since the loss function is positive for all i, we must have $l(b_i, a_i, p_i) = 0$ for all i. This in turn implies that $p_i = m_i$ for all i. Assuming that the model $p_i = f(\Theta, i)$ is injective with respect to Θ, we have a unique parameter $\hat{\Theta}$ that fits our data.

Figure 4.1: Call - implied volatility plot

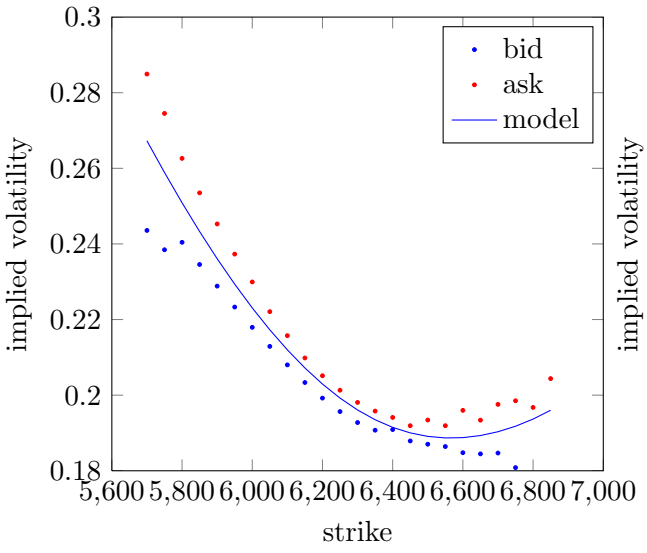
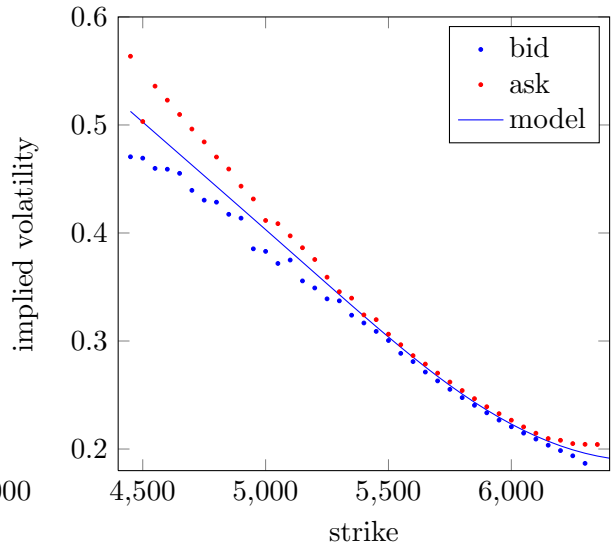


Figure 4.2: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the squared distance to mid loss function, without weight function.

Squared distance to spread

We cannot really distinguish between two fits that are both between the bid-ask spread for all strikes, so therefore it might be an idea to minimize the squared distance to the spread. It is defined as

$$l(b_i, a_i, p_i) := \begin{cases} (p_i - a_i)^2 & \text{if } p_i \geq a_i \\ (p_i - b_i)^2 & \text{if } p_i < b_i \end{cases} . \tag{4.4}$$

Figure 4.3: Call - implied volatility plot

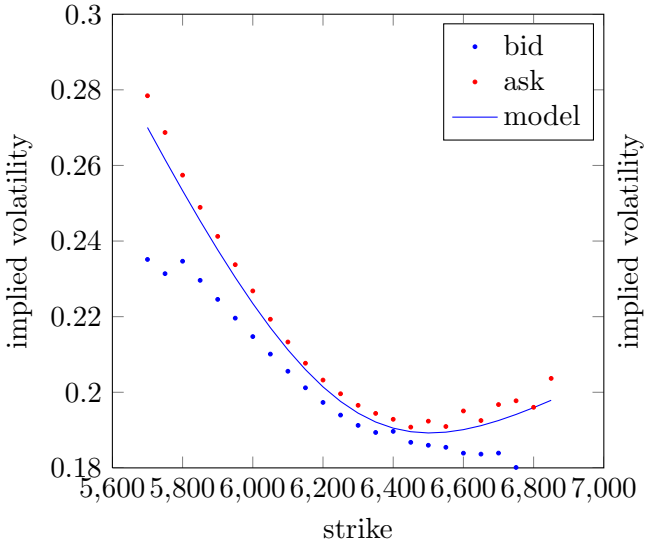
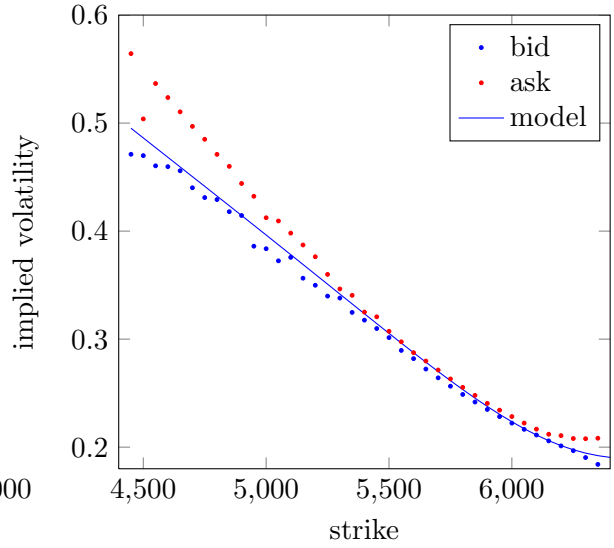


Figure 4.4: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the squared distance to spread loss function, without weight function.

Squared percentage distance to mid

Another approach might be to not focus on the Euro distance to the mid, but instead on the distance as a percentage of the option price itself. We define the squared percentage distance to mid as

$$l(b_i, a_i, p_i) := 100 \left(\frac{p_i - m_i}{m_i} \right)^2 . \tag{4.5}$$

Figure 4.5: Call - implied volatility plot

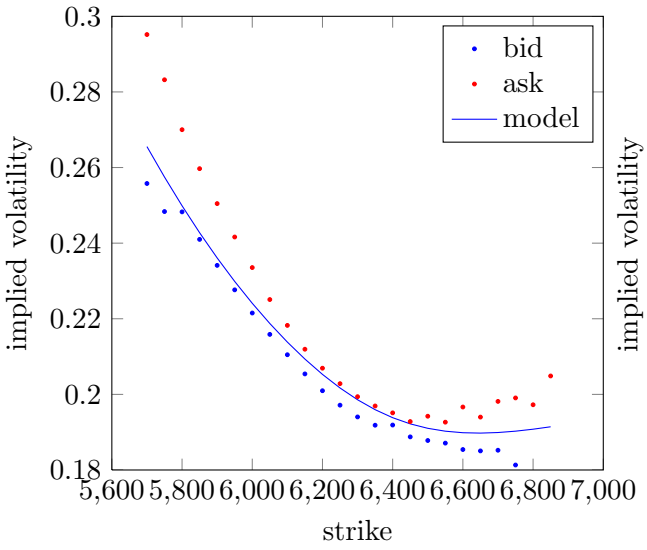
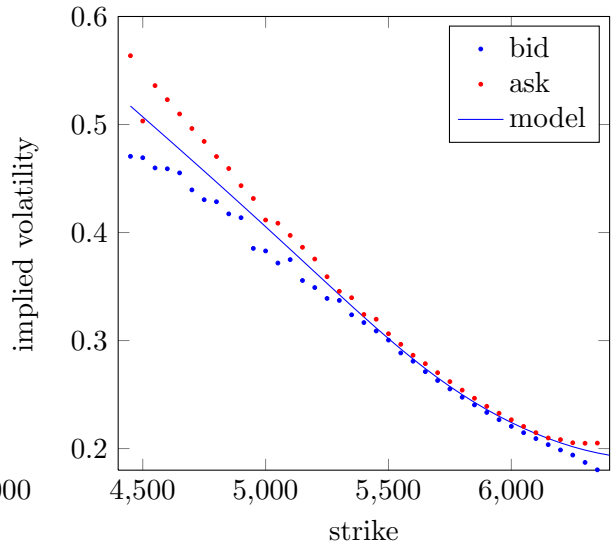


Figure 4.6: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the squared percentage distance to mid loss function, without weight function.

Mid and spread hybrid

We might want to impose an extra penalty to a fit for not being inside the spread at some strike, but if we have to choose, be as close to the mid as possible. Therefore, define a combination between the distance to the mid, and distance to the spread:

$$l(b_i, a_i, p_i) := \begin{cases} (p_i - m_i)^2 + \alpha (p_i - a_i)^2 & \text{if } p_i \geq a_i \\ (p_i - m_i)^2 + \alpha (p_i - b_i)^2 & \text{if } p_i < b_i \end{cases} . \tag{4.6}$$

We work with $\alpha = 3$ in our examples.

Figure 4.7: Call - implied volatility plot

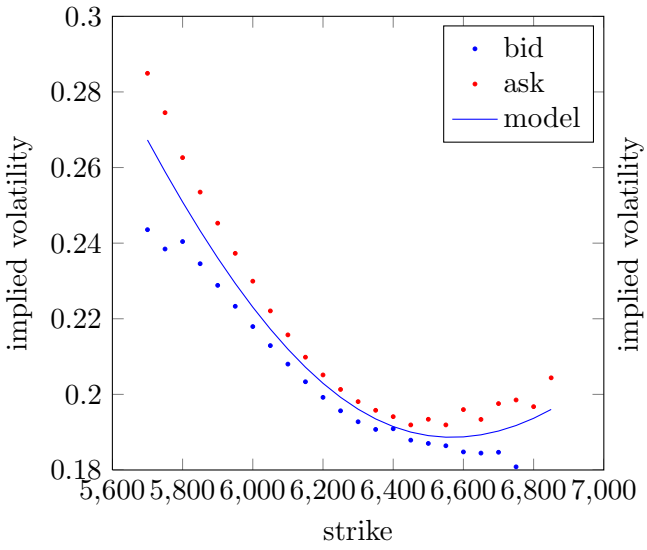
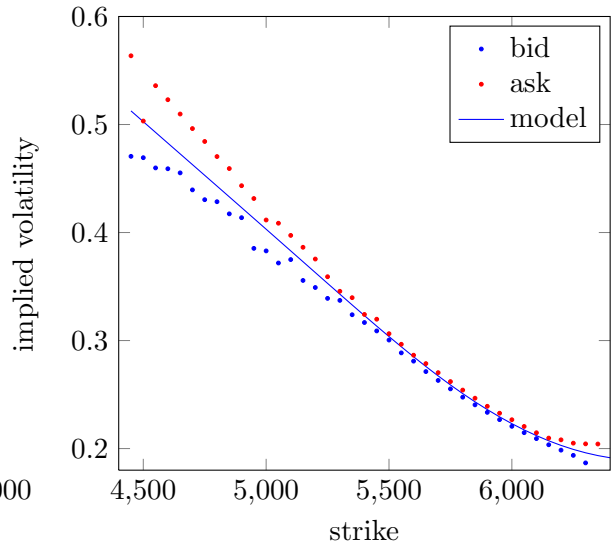


Figure 4.8: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the mid and spread hybrid loss function, without weight function.

Mid and spread hybrid (percentage)

The hybrid loss function can also be stated as a percentage of the mid price:

$$l(b_i, a_i, p_i) := \begin{cases} \left(\frac{p_i - m_i}{m_i}\right)^2 + \alpha \left(\frac{p_i - a_i}{m_i}\right)^2 & \text{if } p_i \geq a_i \\ \left(\frac{p_i - m_i}{m_i}\right)^2 + \alpha \left(\frac{p_i - b_i}{m_i}\right)^2 & \text{if } p_i < b_i \end{cases} . \tag{4.7}$$

Figure 4.9: Call - implied volatility plot

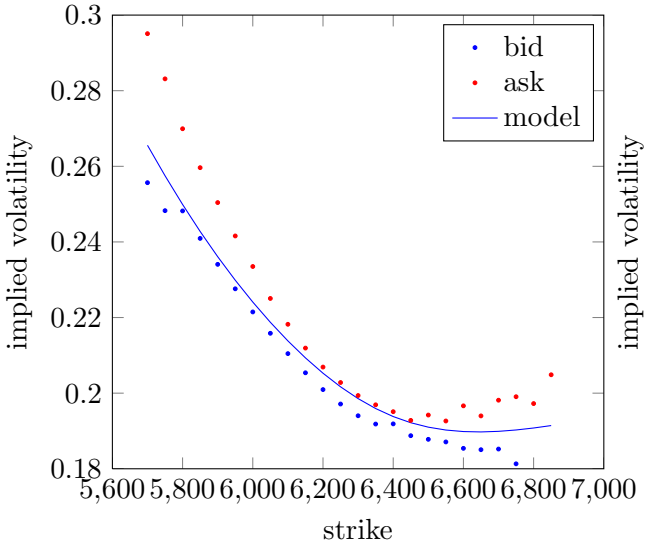
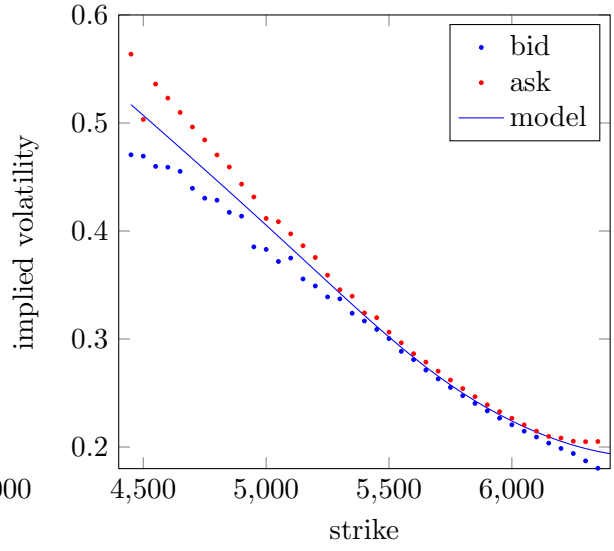


Figure 4.10: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the mid and spread hybrid (percentage) loss function, without weight function.

Absolute distance to mid

In our research we are often interested in the Euro distance of a fit to the mid price. We therefore also define the absolute distance to the mid

$$l(b_i, a_i, p_i) := |p_i - m_i| \tag{4.8}$$

Figure 4.11: Call - implied volatility plot

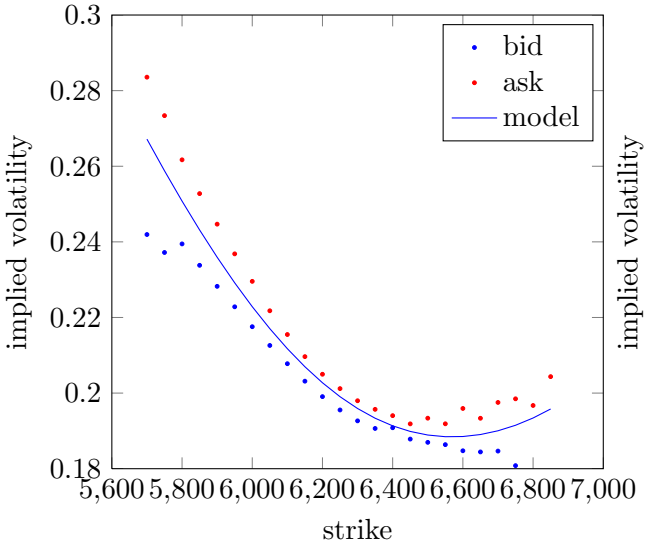
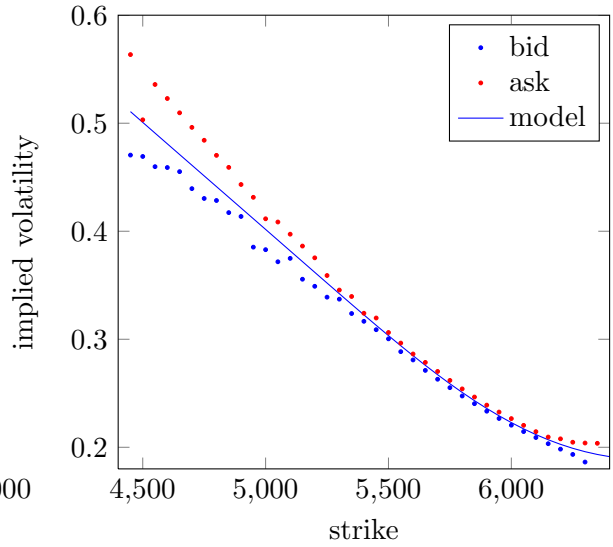


Figure 4.12: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the absolute distance to mid loss function, without weight function.

Conclusion

Through these experiments with various loss functions we have shown that there can exist more than one model that fits the bid-ask spread. Looking at the above graphs, shows us that the main differences are in the wings of the implied volatility curves - that describe the most extreme events. From a practical perspective, one might select the loss function that performs over time in a hedging experiment, or if we model other assets that have a dependence relationship with the modeled asset, we might infer more knowledge to infer knowledge about the 'true' prices. The calibration of the loss function shows the average results for CPU time:

Loss function	**Calibration time (seconds)**
Squared distance to mid	772
Squared distance to spread	7
Squared percentage distance to mid	445
Mid and spread hybrid	865
Mid and spread hybrid (percentage)	472
Absolute distance to mid	174

This shows that there is a lot of variety in the calibration with regard to the loss functions. What surprises is that the squared distance to spread calibrates so fast. We might attribute that to the fact that this loss function has a lot of freedom selecting one particular fit: there might be many fits that fit within the bid-ask spread (and thus have loss value zero) and are equally as attractive to the calibration algorithm. We can see that the algorithm in this case selected a fit that is relatively close to the ask price (in the call option plot). We should not regard this as a good fit.

Altogether, it remains difficult to objectively choose between the various loss functions. The speed of calibration and the easy interpretation supports our choice for the absolute distance to mid as a loss function. We shall use this for the rest of this thesis.

4.4 Weight functions

Besides a loss function we should also weigh observations based on their importance. We will discuss a few examples and comment on them. Without loss of generality, we will define weights $w_i := w(b_i, a_i, p_i)$, such that $\sum_{i=1}^{N} w_i = N$. The spread s_i is defined as $s_i := a_i - b_i$.

No weights

The simplest version is to assign no weights at all:

$$w_i(b_i, a_i, p_i) \quad = \quad 1. \tag{4.9}$$

Inverse spread

We can assign more weight to observations that correspond to a tighter spread (in Euros):

$$w_i(b_i, a_i, p_i) \quad = \quad \frac{1}{s_i}. \tag{4.10}$$

Figure 4.13: Call - implied volatility plot

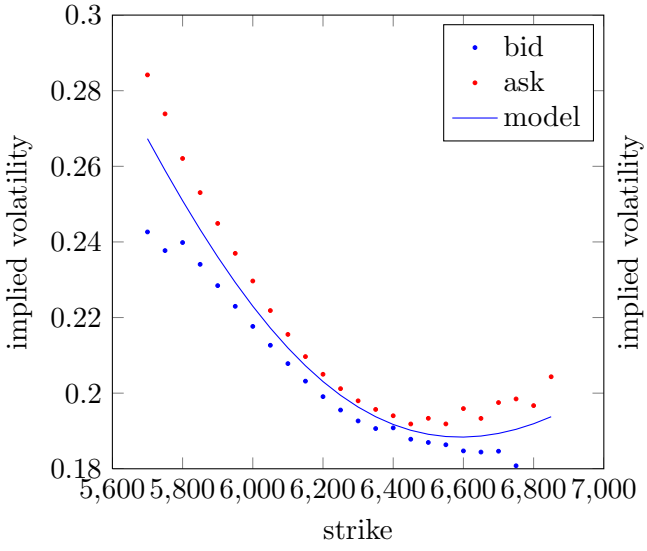
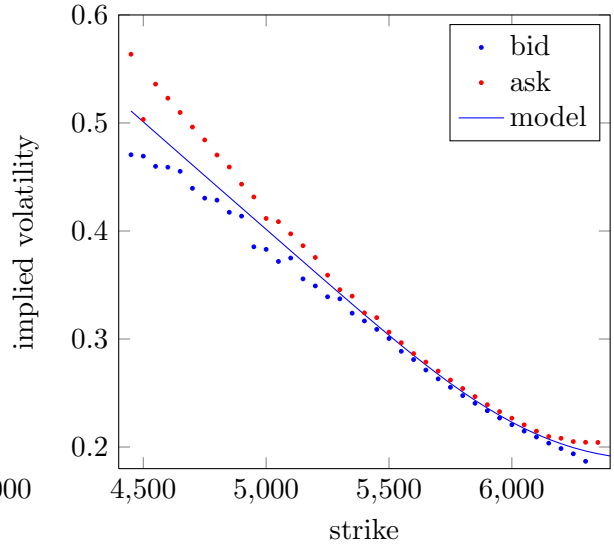


Figure 4.14: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the inverse spread weight function, with the absolute distance to mid loss function.

Maximal spread

The *maximal spread* is defined differently for call and put options. We assume that for both the calls and the puts, the most out of the money options will have the smallest spread. For calls this is the option with the largest strike, and for puts the one with the smallest strike. We then define the weights for the N_C call options with associated spreads $(s_i)_{i=1}^{N_C}$, sorted by their strike, as

$$u_{Ci} := \max(s_i, \dots s_{N_C}) \tag{4.11}$$

and for N_P put options, with associated spreads $(s_i)_{i=1}^{N_P}$ sorted by their strike, we define

$$u_{Pi} := \max(s_1, \dots, s_i). \tag{4.12}$$

We combine the calls and puts through $\mathbf{v} = [\mathbf{u}_C; \mathbf{u}_P]$. Furthermore we set

$$w_i := N \frac{v_i}{\sum_{i=1}^{N} v_i} \tag{4.13}$$

which forces the sum of the weights to be equal to N.

Figure 4.15: Call - implied volatility plot

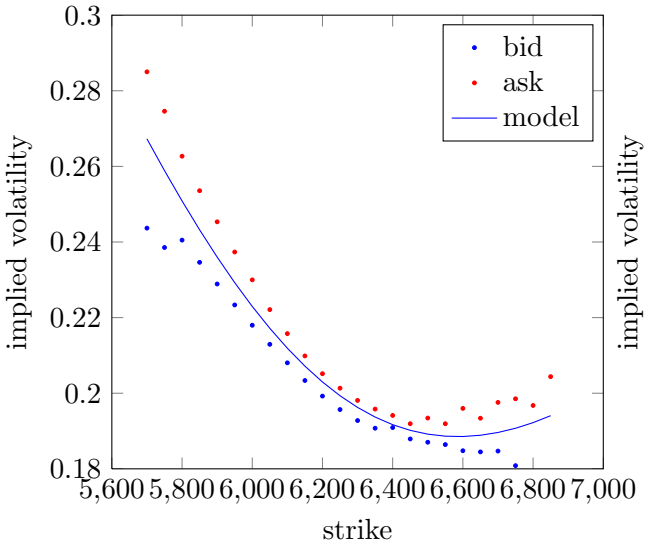
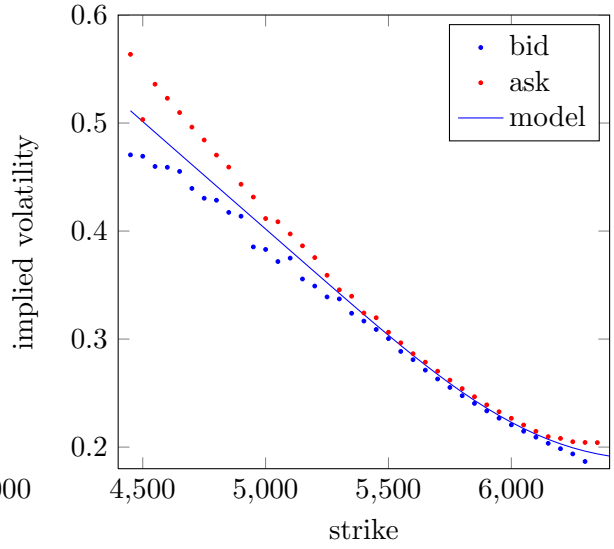


Figure 4.16: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the maximal spread weight function, with the absolute distance to mid loss function.

Remove spread violations (RSV)

It sometimes happens that an observations is clearly out of line since a bid or an ask price is put in near the mid of the adjacent strikes. The spread of that observation is then relatively

small to the spread of the adjacent strikes. We assume the 'real' price in that case not to be near the mid price of that observation, and therefore we would like to discard it. We do this by introducing the variable v_i:

$$v_i := \begin{cases} 0 & \text{if } s_i < s_{i+1} \text{ and } s_i < s_{i-1} \text{ and } s_i < \frac{3}{4}\frac{s_{i-1}+s_{i+1}}{2} \\ 1 & \text{otherwise} \end{cases} \tag{4.14}$$

and then define the weights as

$$w_i := N\frac{v_i}{\sum_{i=1}^{N} v_i} \tag{4.15}$$

to force the sum of the weights to be equal to N.

Figure 4.17: Call - implied volatility plot

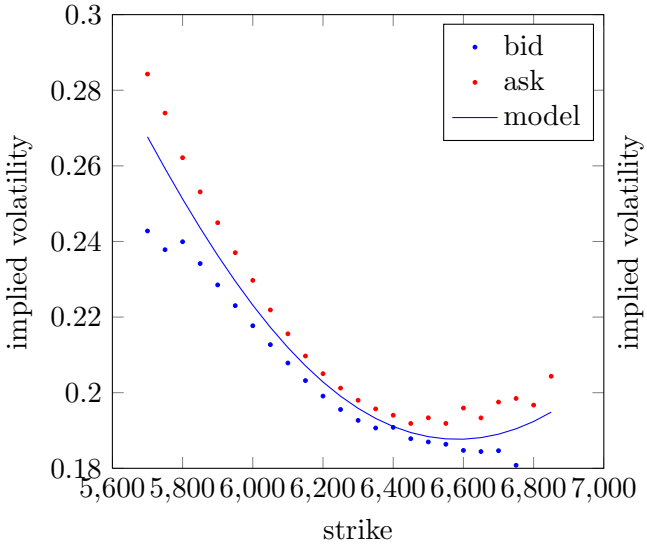
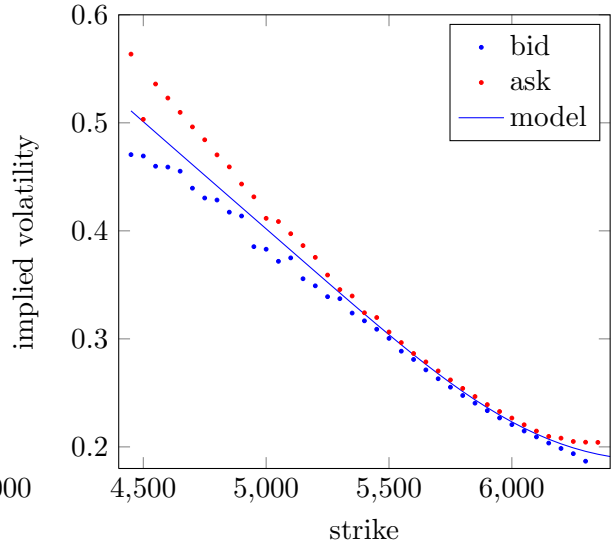


Figure 4.18: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the RSV weight function, with the absolute distance to mid loss function.

Maximal spread and remove spread violations

We can also combine the past two procedures. We first calculate the maximal spread values v_i^{MAX} following equation (4.11) and (4.12). We then calculate v_i^{RSV} from equation (4.14), and define

$$w_i := N\frac{v_i^{\text{MAX}} v_i^{\text{RSV}}}{\sum_{i=1}^{N} v_i^{\text{MAX}} v_i^{\text{RSV}}} \tag{4.16}$$

so that both criteria are taken into account and the sum of the weights is equal to one.

Figure 4.19: Call - implied volatility plot

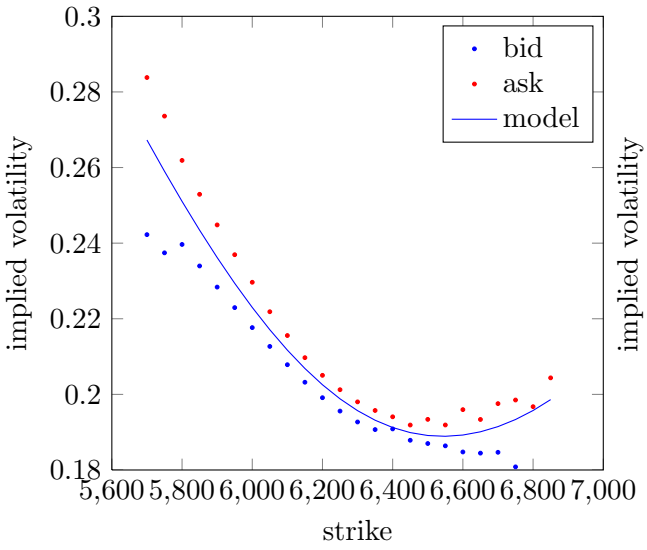
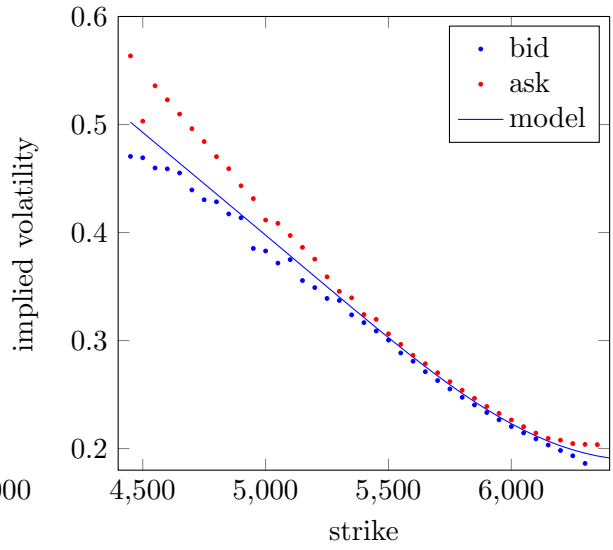


Figure 4.20: Put - implied volatility plot



Above graphs show the Jan-2012 DAX calls and puts, with the Augmented CGMY model (6 parameters) fitted on 4-Jan-2012 at 11:00, using the maximal spread and RSV weight function, with the absolute distance to mid loss function.

Conclusion

Similar to the loss function evaluation, the weight functions also influence the calibration in a significant way. From a trading perspective, determining which observations matter and which don't, is essential in the correct modeling of option prices. The main lesson to be learnt from this is that all these choices heavily influence the selected model, especially in the wings (i.e. the tails of the distribution). Next to this, the calibration of the weight function shows the average results for CPU time:

Weight function	**Calibration time (seconds)**
No weights	-
Inverse spread	97
Maximal spread	56
Remove spread violations	178
Maximal spread and RSV	143

This shows that there is some variety in the calibration with regard to certain weight functions. We choose the inverse spread as a criterion, as we deem it important not to give too much weight to the in the money options - they have wide spreads, and from a trading perspective they behave quite similar to the futures price (the delta is near one), so the prices give less information than the out of the money options. These generally have smaller spreads and therefore the weight criterion matches this choice.

4.5 Calibration algorithms

4.5.1 The Nelder-Mead simplex method

The Nelder-Mead method searches for the minimum of an objective function $L : \mathbb{R}^N \to \mathbb{R}$, using a simplex (a generalization of a triangle in a plane) with $N + 1$ vertices. It calculates the value of the objective function at the vertices, and expands or contracts the simplex until a stopping criterion has been satisfied. The algorithm can be summarized as

1. **Initialization**. We choose an initial simplex $\mathbf{x_1}, \ldots, \mathbf{x_{n+1}}$ with $\mathbf{x_i} \in \mathbb{R}^N$.
2. **Order the points**. Such that $L(\mathbf{x_1}) \leq \ldots \leq L(\mathbf{x_{n+1}})$.
3. **Calculate the center of gravity $\mathbf{x_0}$** for all points except the worst point $\mathbf{x}_{n+1}$, defined as
   $\mathbf{x}_0 := \frac{1}{n} \sum_{i=1}^{N} \mathbf{x}_i$.
4. **Reflection**. From the worst point $\mathbf{x}_{n+1}$ we compute its reflected version $\mathbf{x_r} := \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1})$. If the reflected point is better than $\mathbf{x_n}$, but not better than $\mathbf{x_1}$, then go to step 2. Otherwise proceed to step 5.
5. **Expansion**. If the reflected point $\mathbf{x_r}$ is better than $\mathbf{x_1}$, then compute the expanded point $\mathbf{x_e} := \mathbf{x_0} + \gamma(\mathbf{x_0} - \mathbf{x_{n+1}})$. If the expanded point is better than the reflected point, then set $\mathbf{x_{n+1}} := \mathbf{x_e}$ and go to step 2. Else use the reflected point, $\mathbf{x_{n+1}} := \mathbf{x_r}$ and go to step 2. Otherwise, proceed to step 6.
6. **Contraction**. Compute the contracted point $\mathbf{x_c} := \mathbf{x_0} + \rho(\mathbf{x_0} - \mathbf{x_{n+1}})$. If this point is better than $\mathbf{x_{n+1}}$, then set $\mathbf{x_{n+1}} := \mathbf{x_c}$ and go to step 2. Else, continue to step 7.
7. **Reduction**. For all but the best point $\mathbf{x_1}$, set $\mathbf{x_i} := \mathbf{x_1} + \sigma(\mathbf{x_i} - \mathbf{x_1})$.

The used constants are set to $\alpha = 1, \gamma = 2, \rho = -\frac{1}{2}, \sigma = \frac{1}{2}$.

MATLAB implementation

MATLAB R2012 implements a variant of the above algorithm in the function `fminsearch`. As input it takes some x_0 supplied by the user, and adds to each component 5% of its value to obtain the initial simplex. If the initial value of the component is zero, it adds 0.00025 to the component. It then runs the above routine.

Convergence properties

McKinnon [1998] describes some convergence properties of the Nelder-Mead method. It is shown there that the Nelder-Mead method can sometimes converge to non-stationary points, even though the evaluated function is convex and is three times continuously differentiable. However the method is in practice considered to behave well for most functions, and the fact that it is relatively cheap to use (i.e. does not use derivatives), makes the Nelder-Mead method the method of choice when one does not know much about the function to be minimized.

4.5.2 Gauss-Newton

The Gauss-Newton algorithm considers a function

$$S(\mathbf{p}) \quad = \quad ||y - f(\mathbf{p})||^2 = \sum_{i=1}^{m} (y_i - f(x_i, \mathbf{p}))^2 \tag{4.17}$$

that is the sum of squares of some residual function $r(\mathbf{p}) := y_i - f(x_i, \mathbf{p})$. In our context, $\mathbf{p}$ is the parameter vector of the Lévy process, and each observation i delivers a residual $r_i(\mathbf{p})$ that follows from the chosen loss function. In the case of the squared distance to the mid price, we can consider y_i to be the mid price and $f(x_i, \mathbf{p})$ to be the model price. We can write the first-order Taylor expansion of f around $\mathbf{p}$ in vector notation as

$$f(x_i, \mathbf{p} + \mathbf{\Delta}) \approx f(x_i, \mathbf{p}) + \frac{\partial f(x_i, \mathbf{p})}{\partial \mathbf{p}} \mathbf{\Delta} := f(x_i, \mathbf{p}) + J_i \mathbf{\Delta}. \tag{4.18}$$

where we have defined the Jacobian vector $J_i := \frac{\partial f(x_i, \mathbf{p})}{\partial \mathbf{p}}$. Now the sum of squares function S is approximately

$$S(\mathbf{p} + \Delta) \approx ||y - f(x_i, \mathbf{p}) - J\mathbf{\Delta}||^2 \tag{4.19}$$

where J is the matrix with columns defined by J_i. If we now differentiate the above with respect to the change Δ, and setting this equal to zero, we obtain the *normal equations*

$$(J^T J)\mathbf{\Delta} = J^T(y - f(x_i, \mathbf{p})) = J^T r(\mathbf{p}) \tag{4.20}$$

This is a linear system that we can solve with respect to $\mathbf{\Delta}$. If the matrix $J^T J$ is invertible, we have the solution

$$\mathbf{\Delta} = (J^T J)^{-1} J^T r(\mathbf{p}) \tag{4.21}$$

The case of $J^T J$ being invertible is of course not guaranteed. We can easily see that whenever any column of J is equal to zero, then the determinant of $J^T J$ is zero and it is not invertible. This corresponds to the case when the parameter $\mathbf{p}$ is (locally) optimal in one of its components. When that happens, there can be many solutions to the linear system (4.20). At this point we will not worry about the method to solve this system.

We shall define our algorithm as follows. First we choose an initial parameter $\mathbf{p_0}$. Then for each n we choose $\mathbf{p_n} = \mathbf{p_{n-1}} + \mathbf{\Delta_n}$ where $\mathbf{\Delta_n}$ is found by solving (4.20) for $\mathbf{\Delta_n}$ and $\mathbf{p_n}$.

The problem: approximating the Jacobian

We first repeat the well known finite difference approximation of a derivative. Assume $f \in C^4[a, b]$. The second-order Taylor expansion of f around $\alpha \in [a, b]$ shows

$$f(\alpha + h) = f(\alpha) + hf'(\alpha) + O(h^2). \tag{4.22}$$

From this we can derive the finite difference formula of order 2:

$$f'(\alpha) = \frac{f(\alpha + h) - f(\alpha)}{h} + O(h^2). \tag{4.23}$$

Estimating the Jacobian (or the partial derivatives of the prices with respect to parameters) is difficult since our pricing method is itself an approximation. Suppose now that our pricing function $C(p)$ with p a parameter deviates from the real price $C^*(p)$ by the following

$$C(p) = C^*(p) + \epsilon(p) \tag{4.24}$$

Then rewriting this yields

$$\frac{C(p + h) - C(p)}{h} = \frac{C^*(p + h) - C^*(p)}{h} + \frac{\epsilon(p + h) - \epsilon(p)}{h} \tag{4.25}$$

Then the finite difference formula of order 2 implies that

$$\frac{C(p+h)-C(p)}{h} = C^{*\prime}(p) + O(h^2) + \frac{\epsilon(p+h)-\epsilon(p)}{h} \tag{4.26}$$

But as $h \to 0$, we might have that the term $\frac{\epsilon(p+h)-\epsilon(p)}{h}$ does not converge to zero. In fact it might blow up if the step size h becomes small.

Richardson extrapolation

Suppose that we have an approximation A_h of order k for the derivative. We can write the error in terms of powers of h:

$$f'(\alpha) = A_h + C_k h^k + C_{k+1}h^{k+1} + C_{k+2}h^{k+2} + \dots \tag{4.27}$$

with the $C_i \in \mathbb{R}$ constants. We can use the same approximation for $2h$ instead of h:

$$f'(\alpha) = A_{2h} + C_k(2h)^k + C_{k+1}(2h)^{k+1} + C_{k+2}(2h)^{k+2} + \dots \tag{4.28}$$

Now divide equation (4.28) by 2^k and subtract it from equation (4.27):

$$(1-2^k)f'(\alpha) = A_h - \frac{1}{2^k}A_{2h} - C_{k+1}h^{k+1} - 3C_{k+2}h^{k+2} + \dots \tag{4.29}$$

and thus we have obtained the new formula

$$f'(\alpha) = \frac{A_h - \frac{A_{2h}}{2^k}}{1-2^k} + O(h^{k+1}) \tag{4.30}$$

that is one order higher than the first two formulas. There is of course a tradeoff since the coefficients before the C_k have increased in magnitude, but using this method we can obtain higher-order approximations using relatively larger values of h (which is what we want). We use a repeated Richardson extrapolation in our method to estimate the Jacobian. This allows us to keep the value of h relatively small (in light of the result in equation (4.26).

Line search method

An undesirable property of the Gauss-Newton method in our case is that it can sometimes overshoot: when it is not near a solution, it can move away from the region of acceptable parameter. Therefore we build in a *backtracking line search*, as described in Nocedal and Wright [1999], Chapter 3. At every step of the Gauss-Newton algorithm, we propose a change in the parameter $\mathbf{\Delta}$. In the original algorithm, we define the new parameter by $\mathbf{p}_{k+1} := \mathbf{p}_k + \mathbf{\Delta}$. In the line search, we check whether the *Armijo condition* is satisfied. Initially, set $\alpha = 1$ and check the Armijo condition:

$$f(\mathbf{p}_k + \mathbf{\Delta}) \leq f(\mathbf{p}_k) + c\alpha J^T r(\mathbf{p}_k) \tag{4.31}$$

If this is not satisfied, we set $\alpha = \rho\alpha$, with $\rho < 1$ (in our case we set $\rho = 0.7$), and repeat the procedure. It is proven in Nocedal and Wright [1999] that this procedure ends eventually.

Pros and cons from our experience

When comparing the Nelder-Mead algorithm and Gauss-Newton there are two things to consider. The first is that the Gauss-Newton method has an enormous advantage over the simplex method with regard to the speed of convergence. The use of derivatives is so powerful that in the case of the CGMY process, it reaches a solution within seconds, while the Nelder-Mead method usually takes minutes. However, the errors introduced by the COS method cause the derivative approximation to have limited accuracy. We need to employ smart ways to calculate the derivative without making the step size too small (by using a Romberg extrapolation), but this has limits. We encounter that the Gauss-Newton method gets close to our solution, but only up to a certain accuracy. Another problem is that the Gauss-Newton algorithm is sensitive to the initial guess. If too far away from the minimum, it can 'overshoot' the solution and get stuck in a wrong area (as we deal with a problem with constrained parameters). This is partially solved by introducing a dampening (using a line search).

4.5.3 Conclusions and further research

Having noticed the robustness yet slowness of the Nelder-Mead method and the speed of Gauss-Newton versus its instability, we cannot be really satisfied. For our calibration purposes the robustness is essential, and therefore we use the Nelder-Mead method for any statistics to be gathered. The speed of Gauss-Newton however gives some perspective for further research. One way to go could be to implement the Levenberg-Marquardt ideas of dampening (they use a trust region approach rather than a line search method to keep the stability). The error in the COS method might bound the calculation of derivatives at a small scale, and we might switch to a Nelder-Mead type method when optimizing at the small scale.

Altogether, the Gauss-Newton algorithm gives good results for about 90% of our data. However in the other 10% it fails to converge, and we need to use a Nelder-Mead method to calibrate it reliably. Therefore we have decided to use the best result of both calibration algorithms as the final result of the calibration. Unfortunately this limits the amount of data that we can process as one calibration takes minutes to complete.

4.6 Other calibration issues

4.6.1 Interest rates and obtaining the stock price from the futures price

We need to make some assumption on the risk free rate. It is usually modeled by the discount process D_t that is defined as the stochastic integral

$$D_t := \exp\left(-\int_0^t R_u du\right) \tag{4.32}$$

where R_t is the (stochastic) risk free rate. In our calibration we will assume that the interest rate changes over time - it has a yield curve - but it is deterministic. The risk neutral pricing formula for the value of a contingent claim V_t reads

$$V_t D_t = \mathbb{E}_{\mathbb{Q}}[V_T D_T | \mathcal{F}_t]$$

and when we assume the discount process to be non-stochastic, this changes to

$$V_t = \exp\left(-\int_t^T R_u du\right) \mathbb{E}_{\mathbb{Q}}[V_T | \mathcal{F}_t].$$

For each time of maturity t, the integral $\exp\left(-\int_t^T R_u du\right)$ is just a constant, which we can vary through the calibration. In our research we use the DAX index as underlying instrument for the options. However the index itself (a weighted average of its constituents) is not traded, but the DAX future (a contract that promises to deliver cash equivalent of the index at a certain date) is. The futures price F_t is related to the theoretical index price S_t through the relationship

$$F_t = \mathbb{E}_{\mathbb{Q}}[S_T]$$

We might be surprised that the discount term is not appearing in the above equation. This is due to the fact that a futures contract is continuously settled. See Shreve [2005] for details. Under $\mathbb{Q}$ the discounted stock price $D_t S_t$ is a martingale and thus

$$F_t = \frac{D_T}{D_t} S_t = \exp\left(\int_t^T R_u du\right) S_t$$

in our case of deterministic interest rates. This way we use the futures price to infer knowledge about the theoretical stock price S_t used in the COS method:

$$S_t = \exp\left(-\int_t^T R_u du\right) F_t. \tag{4.33}$$

4.6.2 Considerations for computer implementation: C++ vs. MATLAB

We have used MATLAB for most of our modeling purposes as it provides an easy programming tool with many graphical options for output. It contains many libraries with functions that are useful for a practitioner in applied mathematics. However one of the weaknesses of MATLAB is its poor performance when programming for-loops (i.e. processes that run many times), which is caused by the fact that MATLAB is a scripting language: code is compiled to machine language just before a function is executed. In our calibration we execute an advanced piece of code (the COS method with built in numerical estimation of the characteristic function) many times under different parameters settings. We have therefore chosen to write this code in C++, and connect this to MATLAB using the MEX interface. Using this interface provided for by the authors of MATLAB, one can write a piece of code in C++, compile that to a MEX file (MATLAB's variant of a dynamic link library) and run it from MATLAB as if it were one of its own functions.

Chapter 5

Performance comparison of known Lévy processes

We use Schoutens [2003] as a reference for some well known Lévy processes. We give the definitions and backgrounds of the Variance Gamma, Normal Inverse Gaussian, Meixner, and CGMY processes and the Black Scholes model. We shall give a financial interpretation of the model and its parameters, if that is available. We shall take one cross section of DAX option data, fit the processes to it and discuss the results. Furthermore we shall calibrate all the processes to our entire dataset and discuss the aggregate results.

5.1 The dataset and calibration

Our approach is twofold. First, we have a cross section of DAX option prices on 08-Dec-2011 at 11:15. Using the prices at this time, we shall make some qualitative statements about the performance of models when calibrated to this particular slice. Second, a more extensive analysis is performed with a dataset of option prices of 03-Oct-2011 to 04-Jan-2012, measured every day at 11:15.

We compute the prices with the COS-method with parameters $N = 2^7$, and the automatic integration interval as described in the chapter on the COS method. The measure transform used is the Mean Correcting Martingale Measure. The calibration algorithm is the best result of both Gauss Newton and Nelder-Mead. We calibrate both the option prices and the interest rates (which consist of the rate associated with the maturity of the future, and the one associated with the maturity of the options). The loss function used is the sum of the absolute distances to the mid prices. There is no weighting of observations involved. For details of the mentioned calibration properties we refer to the chapter on calibration.

5.2 The Black-Scholes model (or Geometric Brownian motion)

The Black Scholes model has been introduced in the first chapter. In this model, the logarithm of the stock, $\log(S_t)$ is assumed be governed by a Brownian motion W_t

$$\log(S_t) \quad = \quad (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

where the Brownian motion W_t is a stochastic process such that $W_0 = 0$, having independent and stationary increments, and $X_{t+s} - X_t$ is normally distributed with mean zero and variance

s. It is known that the normal distribution is infinitely divisible, and therefore the Brownian motion is a Lévy process.

The assumed normality of the increments of the log-stock price dates from the time of Bachelier, who modeled the increments of the stock price itself as a Brownian motion. Defining the stock price as the exponent of a Brownian motion guarantees the positivity of the stock price. The characteristic function of the log-stock price is

$$\phi(u) = \exp\left(i(\mu - \frac{1}{2}\sigma^2)u - \frac{1}{2}\sigma^2 u^2\right). \tag{5.1}$$

We are now ready to use the COS-method to calculate prices. The calibration to the single slice shows the result for the first maturity.

Black-Scholes model calibrated to DAX Dec-2011 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.1: Call - absolute distance to mid

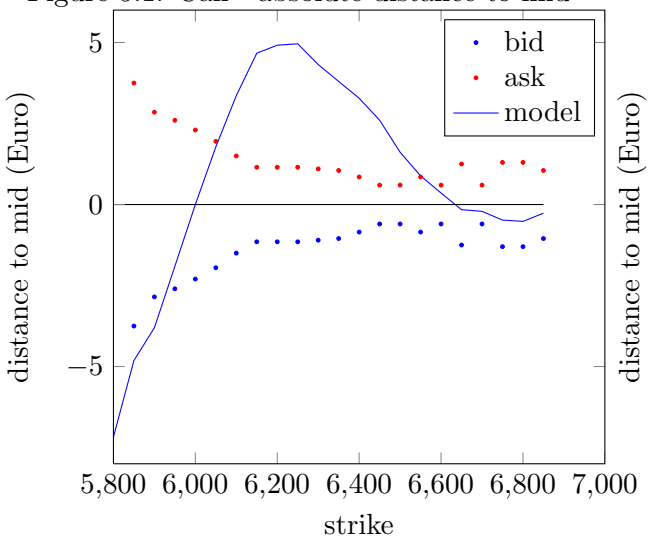


Figure 5.2: Put - distance to mid

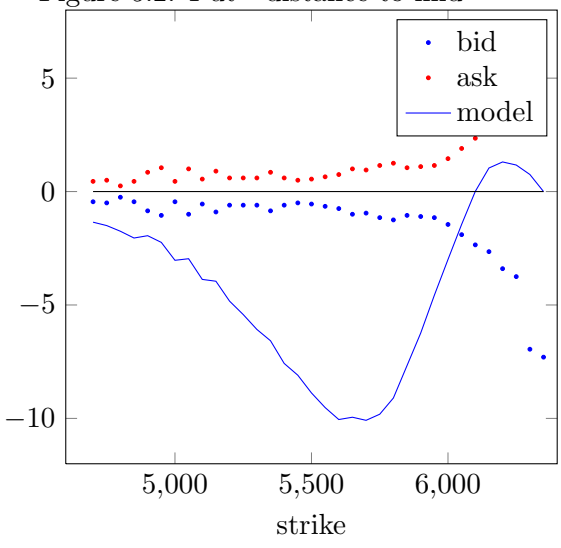


Figure 5.3: Call - implied volatility

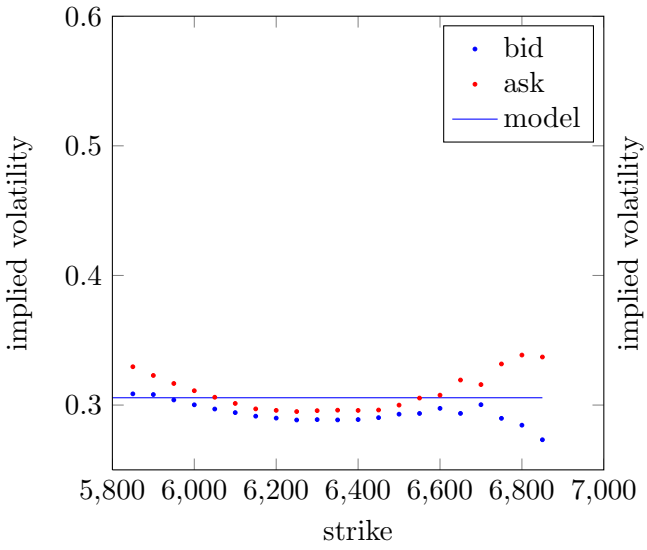
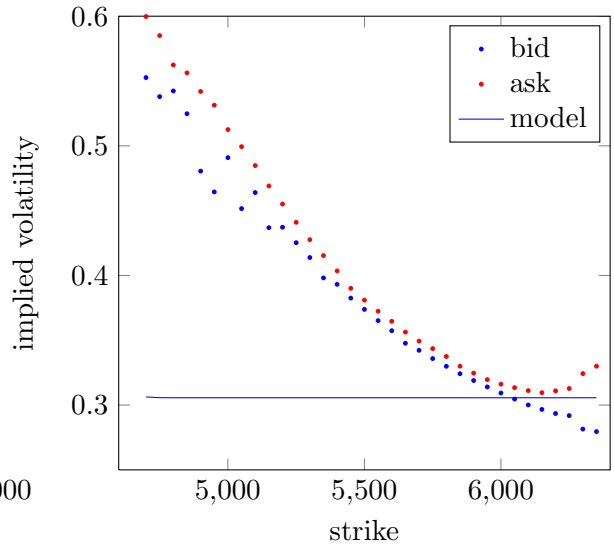


Figure 5.4: Put - implied volatility



The calibrated parameter is $\sigma = 0.31$. We can see that the model is quite far from the observed market data. The model shows a straight line in the implied volatility plot. This follows of course from the definition of the implied volatility, which is based on the Black Scholes model.

5.2.1 Variance Gamma

The Variance Gamma (Madan et al. [1998]) model is defined as a time-changed Brownian motion. If we first consider the model $\theta t + W_t$ with W_t a Brownian motion, we can replace t with a Gamma process G_t (with parameters $a = \frac{1}{v}, b = \frac{1}{v}$), which is an increasing process. The model then becomes

$$L_t = \theta G_t + \sigma W_{G_t} \tag{5.2}$$

Inspiration for the above time change procedure might have come from Monroe's theorem (Monroe [1978]) stating that any càdlàg semimartingale X_t can be written as a time changed Brownian motion W_{T_t} (under some technicalities regarding the setup of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$).

The interpretation of the parameters θ, v, σ in the light of the above definition is as follows. The drift is determined by θ, while σ determines the volatility. The parameter v determines the intensity of the jumps of the Gamma process, and therefore also determines volatility.

Madan et al. [1998] show that the Variance Gamma process can also be defined in terms of its Lévy density

$$\nu(x) = \begin{cases} C \exp(-G|x|)|x|^{-1} & \text{if } x < 0 \\ C \exp(-M|x|)|x|^{-1} & \text{if } x > 0 \end{cases} \tag{5.3}$$

with the parameter transformation

$$C = \frac{1}{v} \tag{5.4}$$

$$G = \left(\sqrt{\frac{1}{4}\theta^2 v^2 + \frac{1}{2}\sigma^2 v} - \frac{1}{2}\theta v\right)^{-1} \tag{5.5}$$

$$M = \left(\sqrt{\frac{1}{4}\theta^2 v^2 + \frac{1}{2}\sigma^2 v} + \frac{1}{2}\theta v\right)^{-1}. \tag{5.6}$$

Later on we will see that this Lévy density is equal to the CGMY process, when the Y-parameter in the CGMY model is equal to zero. The characteristic function of the log-stock process is then derived as

$$\phi(u) = \left(\frac{GM}{GM + (M - G)iu + u^2}\right)^C. \tag{5.7}$$

Calibrating the Variance Gamma model to our slice of data shows the following results.

Variance Gamma model calibrated to DAX Dec-2011 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.5: Call - absolute distance to mid

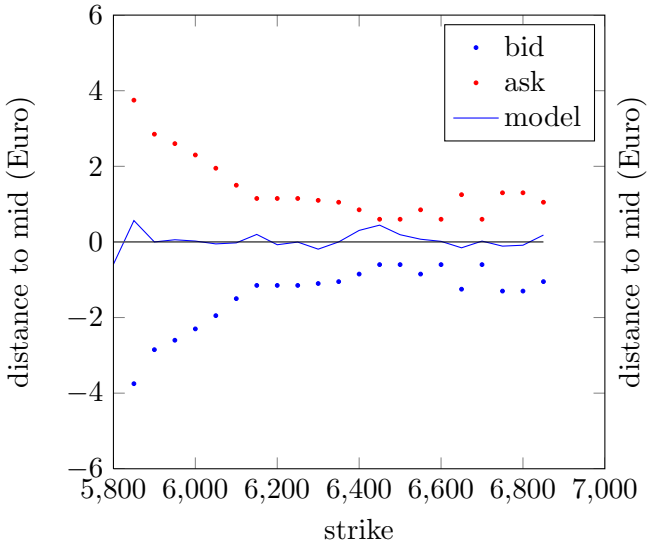


Figure 5.6: Put - distance to mid

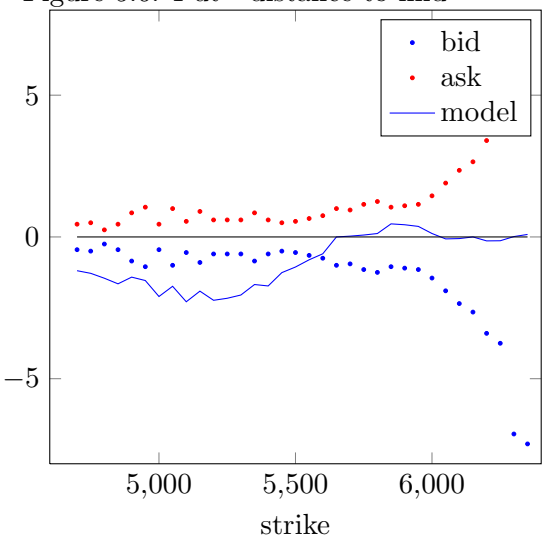


Figure 5.7: Call - implied volatility

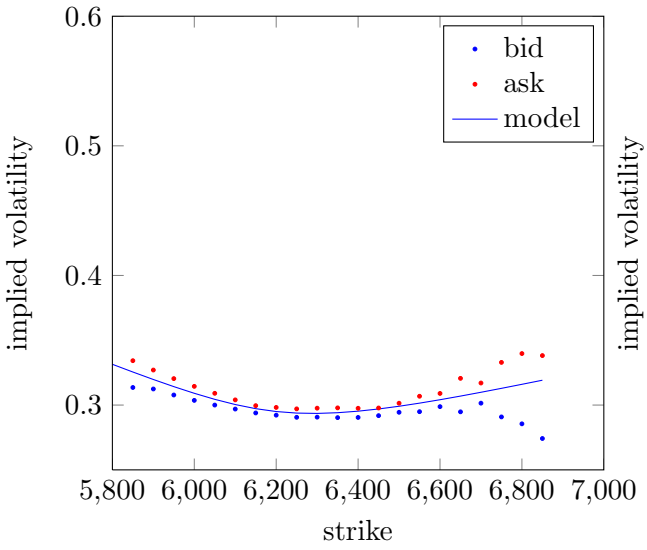
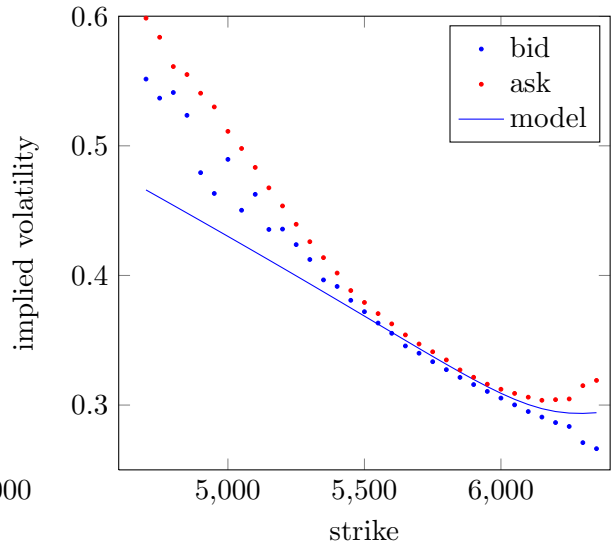


Figure 5.8: Put - implied volatility



The Variance Gamma calibrates with parameters $C = 73.9, G = 31.9, M = 48.4$.

5.2.2 Normal Inverse Gaussian

The NIG process (Barndorff-Nielsen [1997]) can also be related to a time changed Brownian motion. It is defined as a Brownian motion with drift $\beta\delta^2 t + \delta W_t$ time changed with an Inverse Gaussian process I_t with parameters $a = 1, b = \delta\sqrt{\alpha^2 - \beta^2}$, so that it becomes

$$X_t \quad = \quad \beta\delta^2 I_t + \delta W_{I_t}. \tag{5.8}$$

The Inverse Gaussian process is defined as a stopping time

$$I_t \quad = \quad \inf\{s > 0 : \gamma s + W_s = \delta t\} \tag{5.9}$$

Since the process $\gamma s + W_s$ will always hit t before $t + \epsilon$ if $\epsilon > 0$, the Inverse Gaussian process is strictly increasing. Since the time change of a Lévy process by an almost surely increasing Lévy process is still a Lévy process (Cont and Tankov [2004], Theorem 4.2), the NIG process is a Lévy process.

The characteristic function of the NIG model is

$$\phi(u) = \exp\left(-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2})\right). \tag{5.10}$$

A useful practical property of the NIG process is that it has an analytic solution for its density

$$f(x) \quad = \quad \frac{\alpha\delta}{\pi}\exp(\delta\sqrt{\alpha^2 - \beta^2} + \beta x)\frac{K_1(\alpha\sqrt{\delta^2 + x^2})}{\sqrt{\delta^2 + x^2}} \tag{5.11}$$

where K_1 is the modified Bessel function of the third kind with index one. Analytical expressions for associated European option prices are also available. The calibration of NIG shows the following results.

NIG model calibrated to DAX Dec-2011 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.9: Call - absolute distance to mid

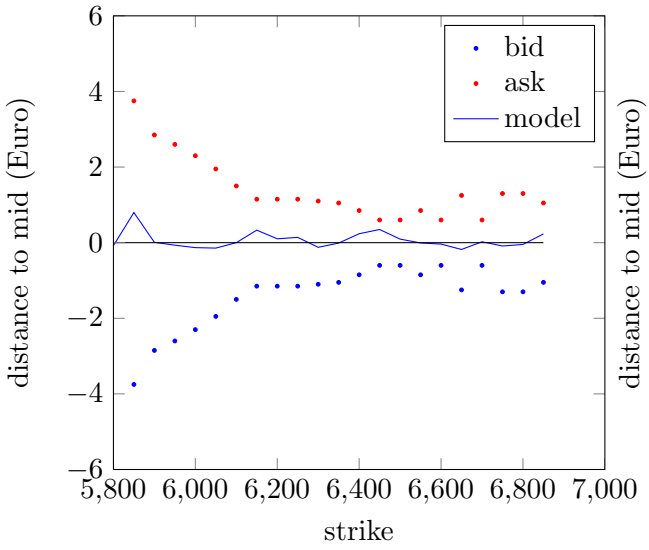


Figure 5.10: Put - distance to mid

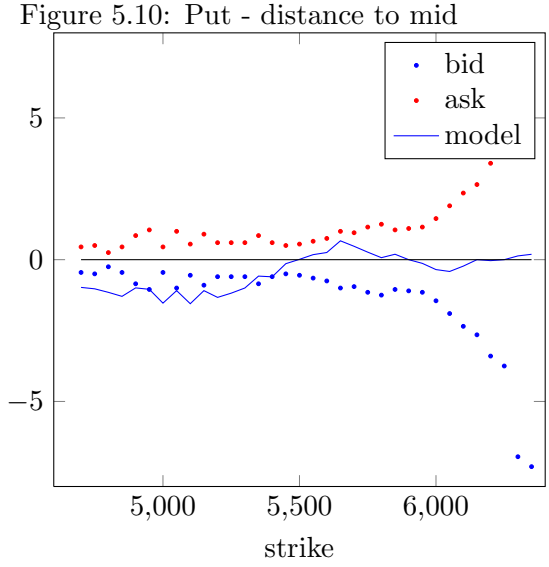


Figure 5.11: Call - implied volatility

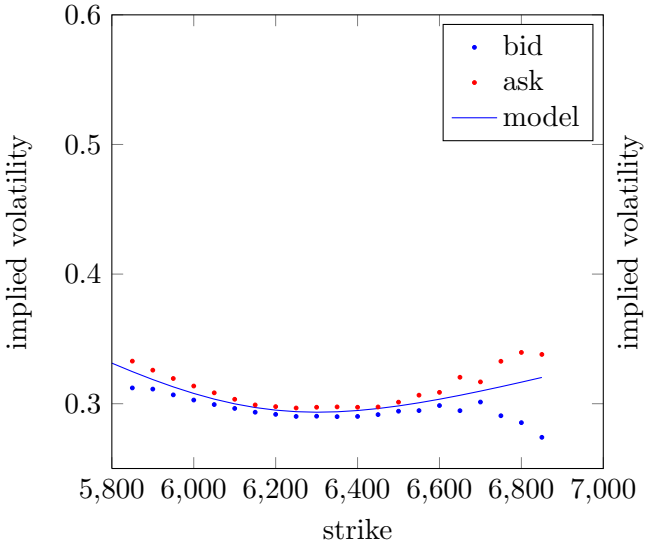
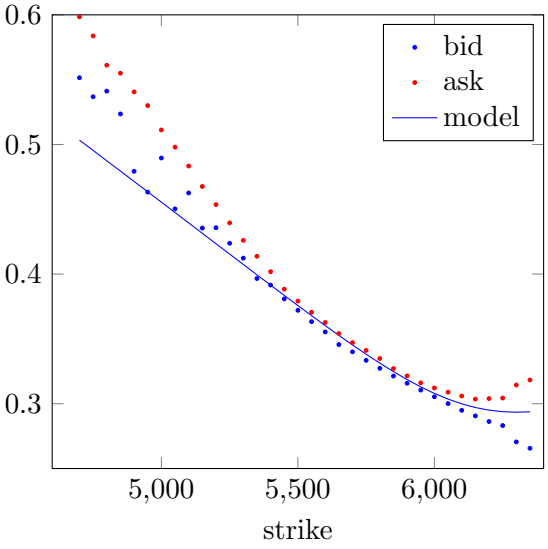


Figure 5.12: Put - implied volatility



NIG calibrated with parameters $\alpha = 25.7, \beta = -7.79, \gamma = 2.36$.

5.2.3 Meixner

The Meixner process comes from Schoutens and Teugels [1998]. Its definition arises in the theory of orthogonal polynomials, with its Lévy density defined as

$$\nu(x) = \delta \frac{\exp(\beta x/\alpha)}{x \sinh(\pi x/\alpha)} \tag{5.12}$$

and a drift component

$$\gamma = \alpha\delta\tan(\beta/2) - 2\delta\int_1^\infty \frac{\sinh(\beta x/\alpha)}{\sinh(\pi x/\alpha)}dx. \tag{5.13}$$

It has no diffusion component. The characteristic function of the Meixner process is defined as

$$\phi_{\text{Meixner}}(u) = \left(\frac{\cos(\beta/2)}{\cosh((\alpha u - i\beta)/2)}\right)^{2\delta}. \tag{5.14}$$

The calibration of Meixner shows the following results.

Meixner model calibrated to DAX Dec-2011 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.13: Call - absolute distance to mid

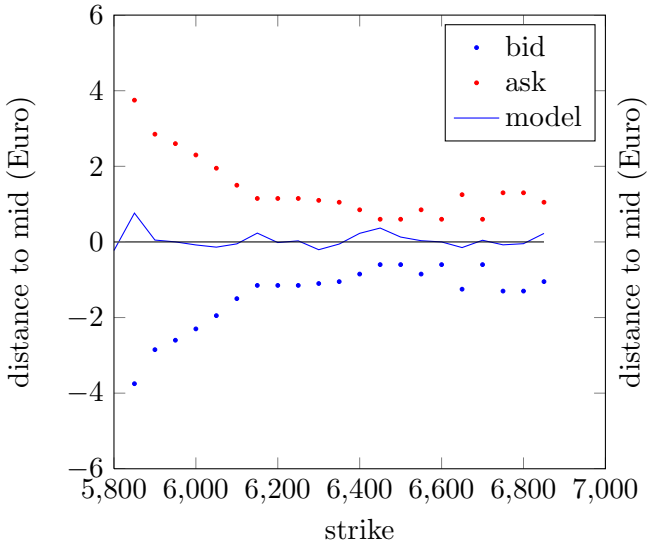


Figure 5.14: Put - distance to mid

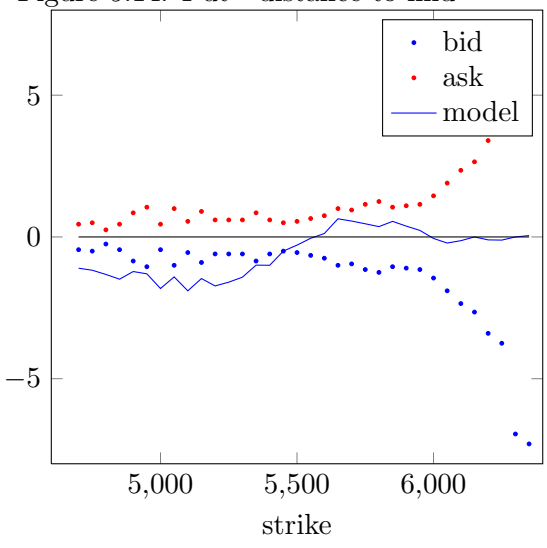


Figure 5.15: Call - implied volatility

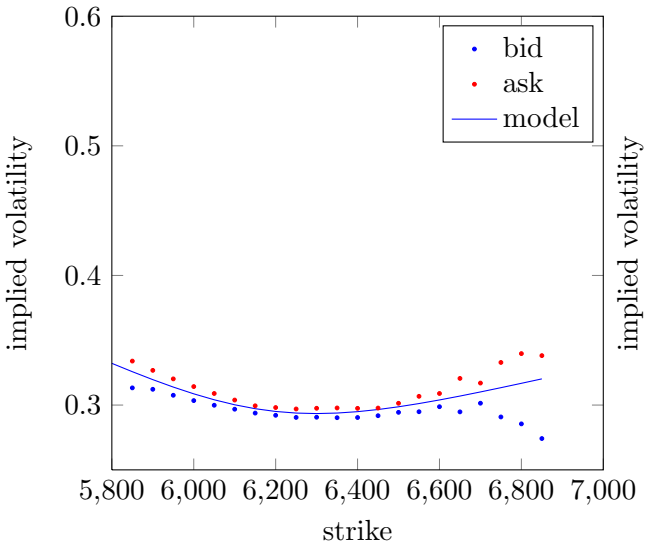
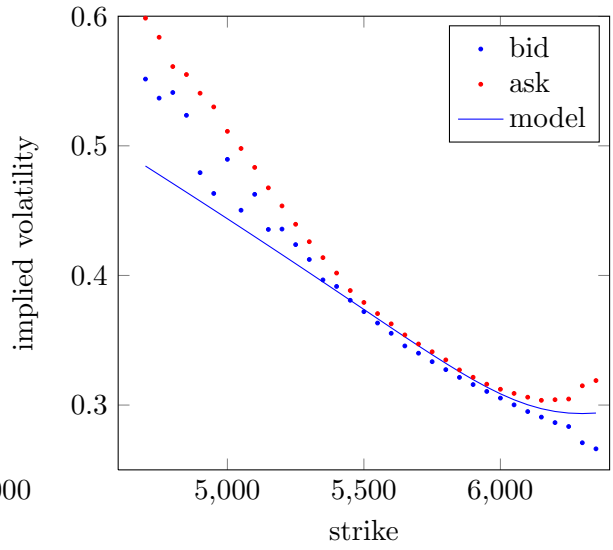


Figure 5.16: Put - implied volatility



Meixner calibrated with parameters $\alpha = 0.09, \beta = -0.73, \delta = 21.77$.

5.2.4 CGMY

The CGMY process (named after the authors of the paper Carr et al. [2002]) is defined by its Lévy density

$$\nu(x) = \begin{cases} C\exp(-G|x|)|x|^{-1-Y} & \text{if } x < 0 \\ C\exp(-M|x|)|x|^{-1-Y} & \text{if } x > 0. \end{cases} \tag{5.15}$$

We see that the parameters G and M influence the exponential decay of the left and right tail. The C parameter determines the total frequency of jumps, while the Y parameter controls mainly the amount of small jumps. A value $Y < 2$ guarantees it to be a Lévy density. If $0 < Y < 1$, the process is of finite variation, and if $1 <= Y < 2$ the process is of infinite variation.

Its characteristic function is derived as

$$\phi_{\text{CGMY}}(u) = \exp\left(C\Gamma(-Y)\left((M-iu)^Y - M^Y + (G+iu)^Y - G^Y\right)\right). \tag{5.16}$$

The calibration for the CGMY process shows the following results.

CGMY model calibrated to DAX Dec-2011 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.17: Call - absolute distance to mid

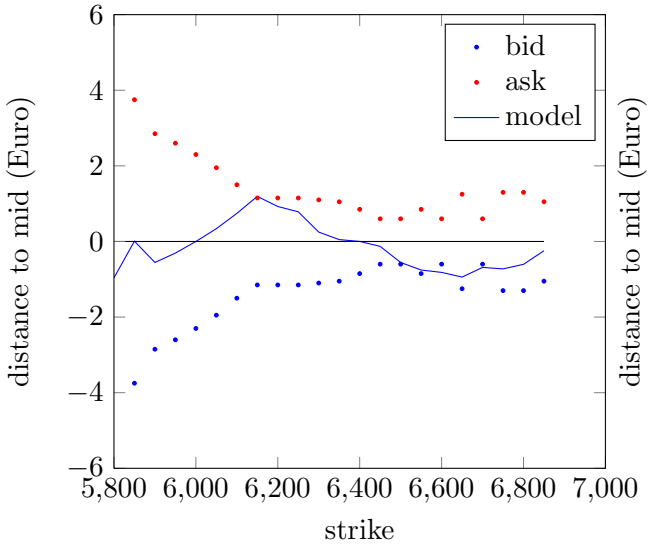


Figure 5.18: Put - distance to mid

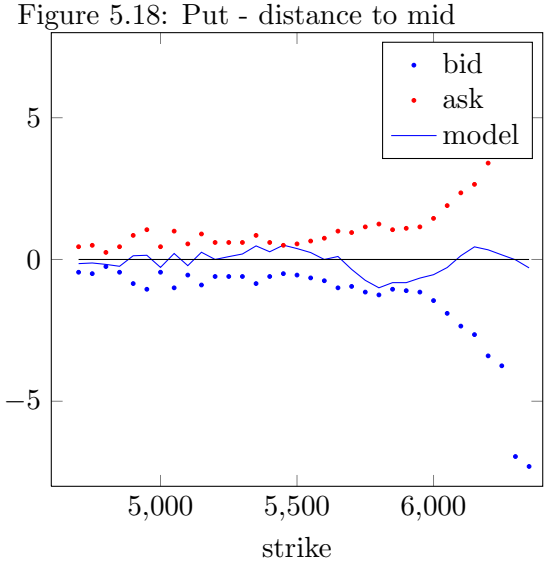


Figure 5.19: Call - implied volatility

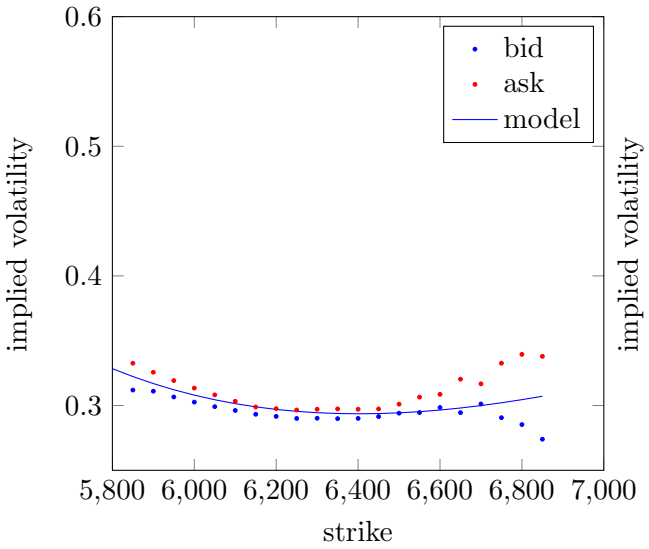
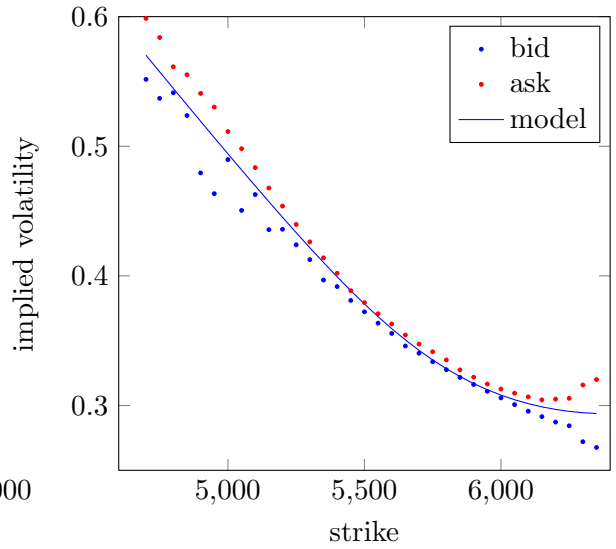


Figure 5.20: Put - implied volatility



The CGMY model calibrated with the parameters $C = 0.09, G = 4.68, M = 23.04, Y = 1.50$.

5.3 Summary of results

The models give the following results for mean distance to the mid prices (MDM), mean absolute distance to the mid price (MADM) and the percentage of observations within the bid-ask spread (%OWS).

Model	MDM	MADM	%OWS	# parameters
Black-Scholes	2.36	3.81	28.6%	1
Variance Gamma	0.50	0.64	67.9%	3
Normal Inverse Gaussian	0.25	0.42	76.8%	3
Meixner	0.32	0.52	73.2%	3
CGMY	0.10	0.35	92.9%	4

In the above results, we have optimized with respect to the mean absolute distance to the mid (MADM). The NIG model and especially the CGMY model perform remarkably well compared to the Black-Scholes model.

This is of course just one particular point in time at which the options and futures data have been recorded. To perform a decent statistical analysis on the performance of the different Lévy models, we fit the models to 3 months of data, of which we fit the model each day at 11:15.

Our methodology here is as follows. For each separate option we infer its Black Scholes implied volatility, and using that (plus an assumption on the interest rate, which we calibrate to the option data using the Put-Call parity) we calculate its Black Scholes delta (the derivative of the option with respect to the underlying). We then create twenty bins of the interval $[-100, 100]$, and place the options in their respective bins. This means that the call options (which all have delta between 0 and 100) will be on the right side, and put options (with delta between -100 and 0) on the left side. Now, we can see how our models perform in different regions of the Black Scholes delta.

Error comparison of models per Black-Scholes delta bin

Figure 5.21: Mean absolute error per Black-Scholes delta bin

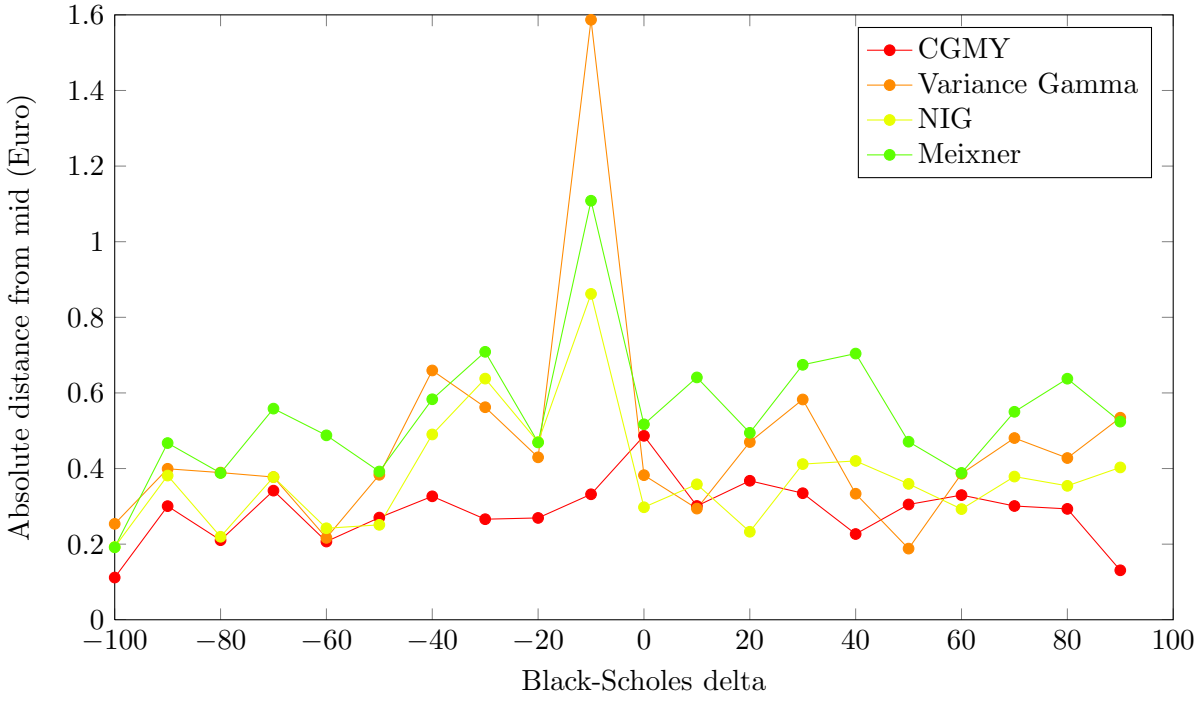
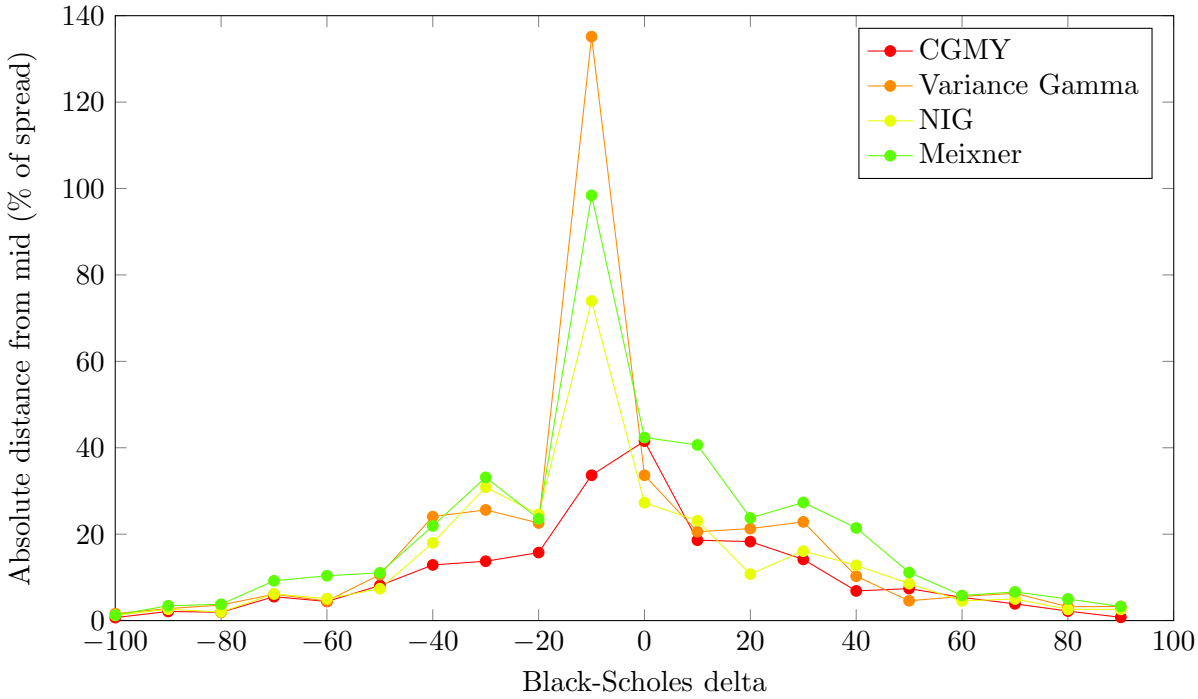


Figure 5.22: Mean absolute error (as % of Euro spread) per Black-Scholes delta bin



We have left the Black-Scholes model out since it performs too bad to be of interest in this comparison. The CGMY model is clearly superior, while having a less good fit in the region (-20,10). Please mind that a dot in the graph at x means in fact the value for the options with delta in the interval $(x, x + 10]$.

5.4 Calibration of multiple maturities

In the following graphs, we have calibrated the CGMY model to the three front month option series, namely the options expiring in December 2011, January 2012, and February 2012. The calibration was done only to the first maturity, as one property of the Lévy model is that the model is time homogeneous, i.e. if it is fixed at one maturity, it is fixed for all maturities.

CGMY model calibrated to DAX Dec-2011, Jan-2012, Feb-2012 Call and Put options on 08-Dec-2011 at 11:15 CET

Figure 5.23: Call prices

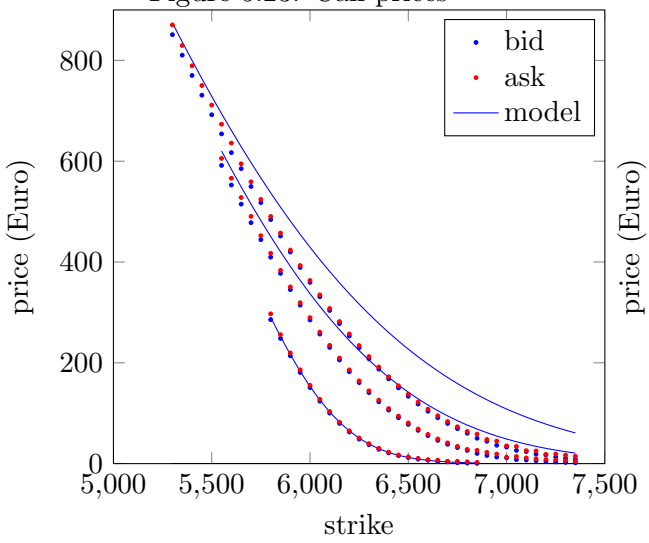


Figure 5.24: Put prices

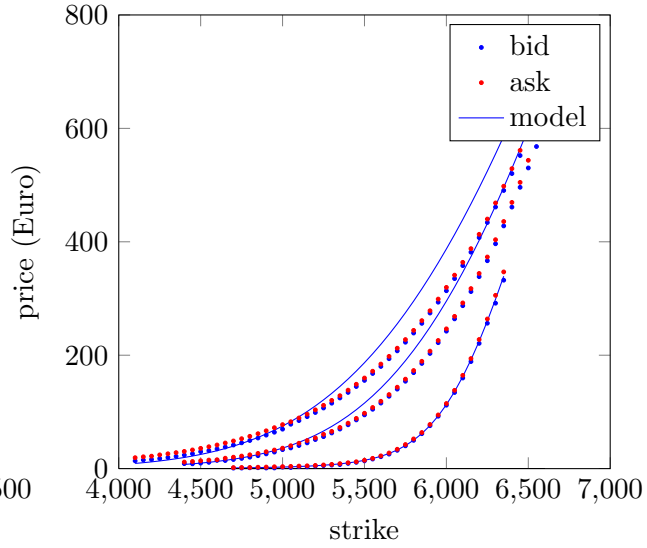


Figure 5.25: Call - implied volatility

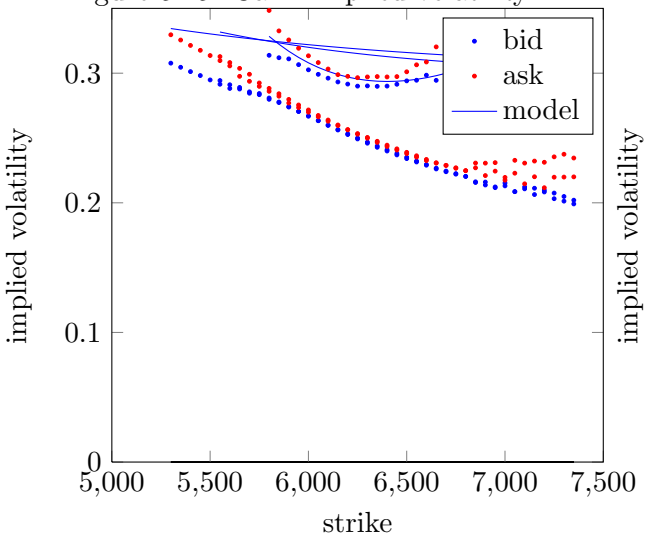
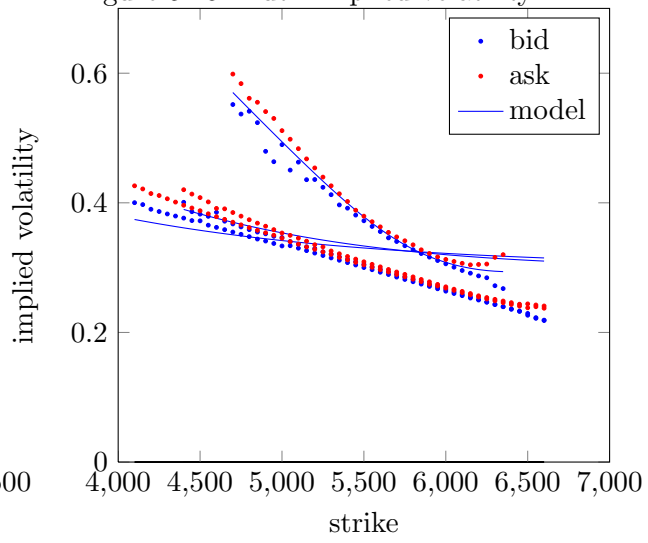


Figure 5.26: Put - implied volatility



The lowest prices are for the December 2011 options, the middle for the January 2012, and the highest for the February 2012 options. What we see is that the fit is reasonable for the first month, but the prices of the options quickly become too high for the later months. It seems that the Lévy model 'lets time run too fast', compared to the 'real' process that corresponds to the market prices of options.

The impossibility of fitting multiple maturities with a Lévy process

The mathematical reason why we are not able to fit all maturities at the same time with a Lévy process is quite simple. The time-homogeneity (or infinite divisibility) assumption determines that the law of the process remains constant over time. This implies in particular that the moments of the process change linearly over time. For instance the variance of a Lévy process L_t is given by

$$\text{Var}(L_t) = t\text{Var}(L_1).$$

This is a very strong assumption that is violated in practice. It tells us that the time structure of the Lévy process is too strict, and we need to relax on that assumption if we wish to model more maturities simultaneously. This can be done using a time change process, as explained in the chapter on time changed Lévy processes.

A second observation is that the implied volatility curve flattens out very fast in the Lévy model. A flat line in the implied volatility surface means that the model returns to log-normality (which could be a desirable property for a model if we consider the central limit theorem). Tankov [2011] notes that in principle the central limit theorem need not apply in general for Lévy processes, and as such the flattening of the implied volatility curve is not per se a consequence of that, but it is rather a consequence of the 'large deviations' principle.

Chapter 6

The Augmented CGMY model

The CGMY model (as introduced in Carr et al. [2002]) fits our market data reasonably well as we have seen in our comparison of known Lévy models. Considering that financial markets are in practice always of a discrete nature (a fixed tick size), its definition as a pure jump process, defined in the Lévy density space, provides an advantage over mixed models (with both a diffusion and a jump component). Fitting individual slices of option data shows us that the CGMY model has problems fitting the out of the money options. These options are mainly influenced by the shape of the tails of the risk neutral probability density. That tail is again heavily influenced by the tails of the Lévy density. We have to be careful, since a change in the Lévy density changes the probability density everywhere, however loosely speaking we could say that a change in the tails of the Lévy density affects the probability density mostly in its tails. This inspires us to propose a model specification as a generalization of the CGMY model to what we call the Augmented CGMY model.

6.1 An extra right tail parameter for the CGMY process

The CGMY model builds a Lévy process by its Lévy density function

$$\nu_{\text{CGMY}}(x) \quad = \quad \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}} & \text{if } x < 0 \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}} & \text{if } x > 0 \end{cases}$$

leading to the characteristic function of the CGMY process

$$\phi_{\text{CGMY}}(u) \quad = \quad \exp\left(tC\Gamma(-Y)\left\{(M-iu)^Y - M^Y + (G+iu)^Y - G^Y\right\}\right).$$

The right tail thus exponentially decays with parameter M while the left tail decays with parameter G. Suppose now that the right tail should decay faster after some point θ_r, say with parameter $M + R$. A Lévy density that carries this behaviour is

$$\nu(x) = \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}} & \text{if } x < 0 \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}} & \text{if } 0 < x < \theta_r \\ C\frac{\exp(-M|x|)\exp(-R|x-\theta|)}{|x|^{1+Y}} & \text{if } x \geq \theta_r \end{cases}$$

The question is whether the neat form of the characteristic function of the CGMY process is preserved by this added parameter. This indeed turns out to be the case, however with the familiar Gamma function replaced by a more general function. The Lévy-Khinchin representation

shows the connection between the Lévy density ν and the characteristic function ϕ:

$$\phi(u) = \exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx\right).$$

The left tail of the integral is the same as in the CGMY process:

$$\begin{aligned}
\int_{-\infty}^{0}(e^{iux}-1)\nu(x)dx &= \int_{-\infty}^{0}(e^{iux}-1)C\frac{\exp(-G|x|)}{|x|^{1+Y}}dx \\
&= C\int_{0}^{\infty}(e^{-iux}-1)\exp(-Gx)x^{-1-Y}dx \\
&= C\int_{0}^{\infty}\exp(-(G+iu)x)x^{-1-Y}dx - C\int_{0}^{\infty}\exp(-Gx)x^{-1-Y}dx \\
&= C(G+iu)^{Y}\int_{0}^{\infty}\exp(-w)w^{-1-Y}dw - CG^{Y}\int_{0}^{\infty}\exp(-w)w^{-1-Y}dw \\
&= C\Gamma(-Y)\left\{(G+iu)^{Y}-G^{Y}\right\}.
\end{aligned}$$

Here we have substituted the *Gamma function*

$$\Gamma(s) := \int_{0}^{\infty}\exp(-w)w^{s-1}dw.$$

Now split the right tail of the integral into two parts:

$$\begin{aligned}
\int_{0}^{\theta_r}(e^{iux}-1)k(x)dx &= \int_{0}^{\theta_r}(e^{iux}-1)C\frac{\exp(-M|x|)}{|x|^{1+Y}}dx \\
&= \int_{0}^{\theta_r}(e^{iux}-1)C\frac{\exp(-Mx)}{x^{1+Y}}dx \\
&= C\int_{0}^{\theta_r}\exp(-(M-iu)x)x^{-1-Y}dx - C\int_{0}^{\theta_r}\exp(-Mx)x^{-1-Y}dx \\
&= C(M-iu)^{Y}\int_{0}^{(M-iu)\theta_r}\exp(-w)w^{-1-Y}dw - CM^{Y}\int_{0}^{M\theta_r}\exp(-w)w^{-1-Y}dw \\
&= C(M-iu)^{Y}\Gamma_{0,(M-iu)\theta_r}(-Y) - CM^{Y}\Gamma_{0,M\theta_r}(-Y)
\end{aligned}$$

where we define the *incomplete Gamma function* as

$$\Gamma_{a,b}(s) := \int_{a}^{b}\exp(-w)w^{s-1}dw.$$

The other part of the right tail is

$$\begin{aligned}
\int_{\theta_r}^{\infty}(e^{iux}-1)k(x)dx &= \int_{\theta_r}^{\infty}(e^{iux}-1)C\frac{\exp(-M|x|)\exp(-H|x-\theta_r|)}{|x|^{1+Y}}dx \\
&= C\int_{\theta_r}^{\infty}\exp(-(M+H+iu)x+H\theta_r))x^{-1-Y}dx \\
&\quad -C\int_{\theta_r}^{\infty}\exp(-(M+H)x+H\theta)x^{-1-Y}dx \\
&= C\exp(H\theta_r)\int_{\theta_r}^{\infty}\exp(-(M+H+iu)x)x^{-1-Y}dx \\
&\quad -C\exp(H\theta_r)\int_{\theta_r}^{\infty}\exp(-(M+H)x)x^{-1-Y}dx \\
&= C\exp(H\theta_r)(M+H+iu)^Y\int_{(M+H+iu)\theta_r}^{\infty}\exp(-w)w^{-1-Y}dw \\
&\quad -C\exp(H\theta_r)(M+H)^Y\int_{(M+H)\theta_r}^{\infty}\exp(-w)w^{-1-Y}dw \\
&= C\exp(H\theta_r)(M+H+iu)^Y\Gamma_{M+H+iu,\infty}(-Y) \\
&\quad -C\exp(H\theta_r)(M+H)^Y\Gamma_{M+H,\infty}(-Y).
\end{aligned}$$

Using the above results, the characteristic function is determined as

$$\begin{aligned}
\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx &= C\Gamma(-Y)\left((G+iu)^Y-G^Y\right)+C(M-iu)^Y\Gamma_{0,(M-iu)\theta_r}(-Y) \\
&\quad -CM^Y\Gamma_{0,M\theta_r}(-Y)+C\exp(H\theta_r)(M+R+iu)^Y\Gamma_{M+R+iu,\infty}(-Y) \\
&\quad -C\exp(H\theta_r)(M+R)^Y\Gamma_{M+R,\infty}(-Y). \qquad (6.1)
\end{aligned}$$

This expression is somewhat more complicated than the characteristic function of the CGMY process. We are lucky in the sense that we can write this analytically, but the introduction of this incomplete Gamma function will turn out to be problematic.

6.1.1 Adding a left tail parameter

We can perform the same exercise to add a left tail parameter, that comes into play below some $\theta_l < 0$. Define the Lévy density of the process as

$$\nu(x) = \begin{cases} C\frac{\exp(-G|x|)}{|x|^{1+Y}} & \text{if } \theta_l < x < 0 \\ C\frac{\exp(-G|x|)\exp(-L|x-\theta_l|)}{|x|^{1+Y}} & \text{if } x \le \theta_l \\ C\frac{\exp(-M|x|)}{|x|^{1+Y}} & \text{if } 0 < x < \theta_r \\ C\frac{\exp(-M|x|)\exp(-R|x-\theta_r|)}{|x|^{1+Y}} & \text{if } x \ge \theta_r \end{cases}.$$

The integral over the left part of the left tail becomes:

$$\begin{aligned}
\int_{-\infty}^{\theta_l}(e^{iux}-1)\nu(x)dx &= C\int_{-\infty}^{\theta_l}(e^{iux}-1)\frac{\exp(-G|x|)\exp(-L|x+\theta_l|)}{|x|^{1+Y}} \\
&= C\int_{-\infty}^{\theta_l}\exp(iux+Gx+L(x+\theta_l))(-x)^{-1-Y}dx \\
&\quad -C\int_{-\infty}^{\theta_l}\exp(Gx+L(x+\theta_l))(-x)^{-1-Y}dx \\
&= Ce^{L\theta_l}\int_{-\infty}^{\theta_l}\exp((iu+G+L)x)(-x)^{-1-Y}dx \\
&\quad -Ce^{L\theta_l}\int_{-\infty}^{\theta_l}\exp((G+L)x)(-x)^{-1-Y}dx \\
&= Ce^{L\theta_l}\int_{\theta_l}^{\infty}\exp(-(iu+G+L)x)x^{-1-Y}dx - Ce^{L\theta_l}\int_{\theta_l}^{\infty}\exp(-(G+L)x)x^{-1-Y}dx \\
&= Ce^{L\theta_l}(iu+G+L)^Y\int_{\theta_l}^{\infty}\exp(-w)w^{-1-Y}dw \\
&\quad -Ce^{L\theta_l}(G+L)^Y\int_{\theta_l}^{\infty}\exp(-w)w^{-1-Y}dx \\
&= Ce^{L\theta_l}(iu+G+L)^Y\Gamma_{(iu+G+L)\theta_l,\infty}(-Y) - Ce^{L\theta_l}(G+L)^Y\Gamma_{(G+L)\theta_l,\infty}(-Y).
\end{aligned}$$

Combining these results with those from the previous section gives us the characteristic function

$$\begin{aligned}
\phi(u) &= Ce^{L\theta_l}(iu+G+L)^Y\Gamma_{(iu+G+L)\theta_l,\infty}(-Y) - Ce^{L\theta_l}(G+L)^Y\Gamma_{(G+L)\theta_l,\infty}(-Y) \\
&\quad +C(M-iu)^Y\Gamma_{0,(M-iu)\theta_r}(-Y) - CM^Y\Gamma_{0,M\theta_r}(-Y) + C\exp(R\theta_r)(M+R+iu)^Y\Gamma_{M+R+iu,\infty}(-Y) \\
&\quad -C\exp(R\theta_r)(M+R)^Y\Gamma_{M+R,\infty}(-Y). \qquad (6.2)
\end{aligned}$$

6.2 General formulation of the Augmented CGMY model class

Our approach has been to extend the Lévy density of the CGMY model. We let the parameters G, M, that control the exponentially decreasing tails, be flexible over the real interval. In general we can write the Lévy density of the Augmented CGMY model as

$$\nu(x) = C|x|^{-1-Y}e^{-G(x)|x|} \qquad (6.3)$$

with $G(x)$ a piecewise constant positive function that attains finitely many different values. So there exists a partition of the real line $-\infty = t_1 < t_2 < \ldots < t_{n-1} < t_n = \infty$, on which $G(x)$ is constant on all intervals (x_i, x_{i+1}). We call the value of $G(x)$ on (x_i, x_{i+1}) to be G_i. The positions x_i will be called θ_i.

In the rest of the thesis we will present the Augmented CGMY model as a version that has one added left tail parameter L and one extra right tail parameter R. This shall prove to be enough for our fitting purposes.

6.3 Implementation

So far our idea for a new Lévy model seems analytically plausible. However, modern software packages such as Matlab do not support the calculation of incomplete Gamma function with

complex-valued parameters. We shall therefore develop our own tool for this. First we shall develop a way to calculate the incomplete Gamma function. Then we shall resort to some numerical integration techniques to calculate the integral $\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx$ directly.

6.3.1 Computing the incomplete Gamma function for complex-valued parameters

For complex z with $\Re(z) > 0$, we have the (nontrivial) identity

$$\Gamma_{0,z}(-Y) + \Gamma_{z,\infty} = \Gamma(-Y).$$

We shall therefore only focus on computing the lower incomplete Gamma function:

$$\gamma_z(x) \quad := \quad \Gamma_{0,z}(x).$$

Equation (6.5.12) of Abramowitz and Stegun [1965] cites the connection between the lower incomplete Gamma function and the *confluent hypergeometric functions*:

$$\Gamma_{0,b}(z) = b^{-1}x^b M(b, 1+b, -x)$$

where M is the Kummer M-function, a particular hypergeometric function. To approximate this function, we use the series expansion of the Kummer M-function. Equation (13.1.2) from Abramowitz and Stegun [1965] reads:

$$M(a,b,z) \quad = \quad \sum_{n=0}^{\infty} \frac{a_n z^n}{b_n n!}$$

where

$$\begin{aligned} a_n &= a(a+1)(a+2)\cdots(a+n-1) \\ b_n &= b(b+1)(b+2)\cdots(b+n-1). \end{aligned}$$

Our approximation truncates this series at some integer N:

$$M_N(a,b,z) \quad := \quad \sum_{n=0}^{N} \frac{a_n z^n}{b_n n!}$$

and uses this as a proxy for $M(a,b,z)$ in further calculations. The difficulty with the above approximation is that it suffers from roundoff errors. We see that the numerators and denominators can reach very large values; this introduces a roundoff error that can blow up as N becomes large.

6.3.2 Gaussian quadrature numerical integration

Suppose we wish to numerically integrate an integral $\int_a^b g(t)dt$. A Gaussian quadrature method hopes that we can write $g(t) = w(t)f(t)$ with w a known function and f a polynomial. We then try to estimate this integral by

$$\int_a^b g(t)dt \quad = \quad \int_a^b w(t)f(t)dt \approx \sum_{i=1}^{N} w_i f(x_i).$$

The weights w_i and abscissas x_i are to be determined such that the approximation is exact for f in some set of polynomials. References to this methodology are Gautschi [1968] and

Golub and Welsch [1967]. The weights and the abscissas are set up precisely such that the formula is exact for all polynomials of order $2N + 1$.

The precise calculation of the weights and abscissas can be quite involved, depending on the chosen weight function w. We consider this outside of the scope of this thesis. Since the subject has been so extensively researched, a wide range of tools are available to implement it in a software package without any hassle.

One important note that needs to be made is that this approximation depends on if f can be approximated well by a polynomial. If the function f is for instance an exponential function, we might suffer problems with oscillations of the polynomials.

Generalized Gauss-Hermite quadrature

Define the weight function as

$$w(x) \quad = \quad x^{\alpha} e^{-bx^2}.$$

This is the weight function associated with the Gauss-Hermite quadrature. Suppose now that we wanted to evaluate the integral

$$I \quad := \quad \int_0^{\infty} x^5 e^{-2x^2} dx.$$

The obvious strategy would be to use the Generalized Gauss-Hermite quadrature with $\alpha = 2$. Then set $f(x) = x^5$ and calculate the sum

$$\hat{I} \quad := \quad \sum_{i=1}^{N} w_i f(x_i)$$

with the weights w_i and abscissas x_i prescribed by the Gauss-Hermite quadrature. As x^5 is a polynomial, this approximation will be exact for any $N \geq 2$. This is of course a very convincing example why to use a Gaussian quadrature method. This particular quadrature is interesting because the form e^{-bx^2} resembles the structure of the normal probability density.

Generalized Gauss-Laguerre quadrature

The Generalized Gauss-Laguerre quadrature is based on the weight function

$$w(x) \quad := \quad x^{\alpha} e^{-bx}$$

with the restriction $\alpha > -1$.

This quadrature is of interest since the weight function has the same form as the Lévy density of the CGMY model. For the above two quadratures, we can find packages to calculate the weights and abscissas for MATLAB and other languages on the website of John Burkardt at `http://people.sc.fsu.edu/~{}jburkardt/`.

Problems with oscillations

A fundamental problem with this type of Gaussian quadrature is that convergence is established on the hypothesis that the integrand can be well approximated by some polynomial of arbitrary

order. In our practice, we wish to integrate Lévy density functions that are often exponentially decreasing (such as the CGMY model). We need a very large order of the polynomials to achieve a low error, and the error in the resulting integrand causes oscillations in the resulting probability density (when integrated from the Levy density to the characteristic function and then to the density).

6.3.3 Double Exponential numerical integration

The Double Exponential integration technique (sometimes called the Tanh-Sinh quadrature) was developed in Takahasi and Mori [1974]. It approximates an integral $I = \int f(x)dx$ by a change of variables, such that the resulting integrand decays fast. The main idea is that a change of variables by a continuous function $g(x)$

$$\int_{\Omega} f(x)dx = \int_{\hat{\Omega}} f(g(x))g'(x)dx$$

can change the integrand such that it becomes of the form

$$f(g(x))g'(x) = O((\exp(\exp(-\lambda|x|))) \quad \text{as } |x| \to \infty. \tag{6.4}$$

It has been shown in Takahasi and Mori [1974] that such an integrand can be evaluated fast using a simple trapezoidal formula

$$I_h := h \sum_{-N}^{N} f(g(kh))g'(kh)$$

with step size h, and an equally spaced grid, and N large enough. In fact it was proven by Sugihara [1986] that the trapezoidal formula is optimal (uses the least number of function evaluations) in the space of doubly exponentially decaying functions.

This approach has changed the problem of finding a suitable integration formula (such as the trapezoidal formula) into finding a change of variable formula, depending on the function to be integrated. It is well known that the Tanh-Sinh formula, as a first proposal in Takahasi and Mori [1974],

$$g(x) = \tanh\left(\frac{1}{2}\pi \sinh(x)\right) \tag{6.5}$$

performs well for many regular functions, but not so well for oscillatory integrands (we need a large number of evaluations to compensate for this). In our problem we work with oscillatory integrands as we are interested in the calculation of Fourier integrals $\int e^{iux}\nu(x)dx$ with $\nu(x)$ fast decreasing. Ooura [2005] deals with this problem and provides a way to compute Fourier-type integrals using a suitable transformation.

An implementation of the original Double Exponential formula in C++ by John Cook is available on the Code Project website available at `http://www.codeproject.com/Articles/31550/Fast-Numerical` We have used a slightly adapted version of this software to calculate the results in this thesis.

An interesting read on the development of the Double Exponential formula and its optimality is Mori [2005]. It is a promising method for integrating Lévy densities, as it has no problem with a possible singularity at zero. That together with its ease of implementation, it could be a

better choice than the Fast Fourier Transform.

The actual implementation of the Double Exponential integration for our model is done as follows. We need to calculate the characteristic function ϕ_t

$$\phi_t(u) = \exp\left(t\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx\right).$$

The integral can be split up into a real and an imaginary part:

$$\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx = \int_{\mathbb{R}\setminus\{0\}}(\cos(ux)-1)\nu(x)dx + i\int_{\mathbb{R}\setminus\{0\}}\sin(ux)\nu(x)dx$$

and we can split up the negative and positive part of the real line and perform a change of variables to obtain four real integrals from zero to infinity:

$$\begin{aligned}\int_{\mathbb{R}\setminus\{0\}}(e^{iux}-1)\nu(x)dx &= \int_0^\infty(\cos(ux)-1)\nu(x)dx + \int_0^\infty(\cos(ux)-1)\nu(-x)dx \\ &\quad +i\int_0^\infty\sin(ux)\nu(x)dx - i\int_0^\infty\sin(ux)\nu(-x)dx.\end{aligned}$$

These integrals are calculated in this thesis using the implementation by John Cook as stated above.

6.4 Moments calculation

The Lévy density of the CGMY process is

$$\nu(x) = \begin{cases} C|x|^{-1-Y}e^{-G|x|} & \text{if } x<0 \\ C|x|^{-1-Y}e^{-M|x|} & \text{if } x>0. \end{cases}$$

Let us calculate the n-th moment of this density:

$$\begin{aligned} m_n &= \int x^n\nu(x)dx = C\left(\int_{-\infty}^0 x^n(-x)^{-1-Y}e^{Gx}dx + \int_0^\infty x^n x^{-1-Y}e^{-Mx}dx\right) \\ &= C\left(\int_0^\infty(-x)^n x^{n-1-Y}e^{Gx}dx + \int_0^\infty x^n x^{-1-Y}e^{-Mx}dx\right) \\ &= \begin{cases} C\left(\int_0^\infty x^{n-1-Y}e^{Gx}dx + \int_0^\infty x^{n-1-Y}e^{-Mx}dx\right) & \text{if } n \text{ even} \\ C\left(-\int_0^\infty x^{n-1-Y}e^{Gx}dx + \int_0^\infty x^n x^{-1-Y}e^{-Mx}dx\right) & \text{if } n \text{ odd} \end{cases} \\ &= \begin{cases} C\left(G^{Y-n}\int_0^\infty x^{(n-Y)-1}e^x dx + M^{Y-n}\int_0^\infty x^{(n-Y)-1}e^{-x}dx\right) & \text{if } n \text{ even} \\ C\left(-G^{Y-n}\int_0^\infty x^{(n-Y)-1}e^x dx + M^{Y-n}\int_0^\infty x^n x^{(n-Y)-1}e^{-x}dx\right) & \text{if } n \text{ odd} \end{cases} \\ &= \begin{cases} C\left(G^{Y-n}\Gamma(n-Y) + M^{Y-n}\Gamma(n-Y)\right) & \text{if } n \text{ even} \\ C\left(-G^{Y-n}\Gamma(n-Y) + M^{Y-n}\Gamma(n-Y)\right) & \text{if } n \text{ odd} \end{cases}. \end{aligned}$$

6.4.1 Moments of the Augmented CGMY model

Start with the Lévy density of the Augmented CGMY model:

$$\nu(x) = C|x|^{-1-Y}e^{-G(x)|x|} \tag{6.6}$$

with $G(x)$ a piecewise constant function that attains finitely many different values G_i. Determine the n-th moment of ν:

$$\begin{aligned}
m_n &= \int_{-\infty}^{\infty} x^n \nu(x)dx = C\int_{-\infty}^{\infty} x^n |x|^{-1-Y} e^{-G(x)|x|}dx \\
&= C\left(\int_{-\infty}^{0} x^n(-x)^{-1-Y}e^{G(x)x}dx + \int_0^{\infty} x^n x^{-1-Y}e^{-G(x)x}dx\right) \\
&= C\left(\int_{-\infty}^{0} (-x)^n x^{-1-Y}e^{-G(x)x}dx + \int_0^{\infty} x^n x^{-1-Y}e^{-G(x)x}dx\right) \\
&= C\left(\sum_{i=1}^{M-1}\int_{-t_i}^{t_{i+1}} (-x)^n x^{-1-Y}e^{-G(x)x}dx + \sum_{i=M}^{n-1}\int_{x_i}^{x_{i+1}} x^n x^{-1-Y}e^{-G(x)x}dx\right) \\
&= \begin{cases} C\left(\sum_{i=1}^{M-1}\int_{-t_i}^{-t_{i+1}} x^n x^{-1-Y}e^{-G(x)x}dx + \sum_{i=M}^{n-1}\int_{t_i}^{t_{i+1}} x^n x^{-1-Y}e^{-G(x)x}dx\right) & \text{if } n \text{ even} \\ C\left(-\sum_{i=1}^{M-1}\int_{-t_i}^{-t_{i+1}} x^n x^{-1-Y}e^{-G(x)x}dx + \sum_{i=M}^{n-1}\int_{t_i}^{t_{i+1}} x^n x^{-1-Y}e^{-G(x)x}dx\right) & \text{if } n \text{ odd} \end{cases} \\
&= \begin{cases} C\left(\sum_{i=1}^{M-1} G_i^{Y-n}\int_{-t_i}^{-t_{i+1}} x^{(n-Y)-1}e^{-x}dx + \sum_{i=M}^{n-1} G_i^{Y-n}\int_{t_i}^{t_{i+1}} x^{(n-Y)-1}e^{-x}dx\right) & \text{if } n \text{ even} \\ C\left(-\sum_{i=1}^{M-1} G_i^{Y-n}\int_{-t_i}^{-t_{i+1}} x^{(n-Y)-1}e^{-x}dx + \sum_{i=M}^{n-1} G_i^{Y-n}\int_{t_i}^{t_{i+1}} x^{(n-Y)-1}e^{-x}dx\right) & \text{if } n \text{ odd} \end{cases} \\
&= \begin{cases} C\left(\sum_{i=1}^{M-1} G_i^{Y-n}\Gamma_{-t_i,-t_{i+1}}(n-Y) + \sum_{i=M}^{n-1} G_i^{Y-n}\Gamma_{t_i,t_{i+1}}(n-Y)\right) & \text{if } n \text{ even} \\ C\left(-\sum_{i=1}^{M-1} G_i^{Y-n}\Gamma_{-t_i,-t_{i+1}}(n-Y) + \sum_{i=M}^{n-1} G_i^{Y-n}\Gamma_{t_i,t_{i+1}}(n-Y)\right) & \text{if } n \text{ odd} \end{cases}
\end{aligned}$$

where $\Gamma_{a,b}(s) := \int_a^b x^{s-1}e^{-x}dx$ is the incomplete Gamma function.

6.5 Change of measure

To move between the real world probability measure $\mathbb{P}$ and the risk neutral measure $\mathbb{Q}$, a desirable property would be to find a measure change that preserves the model class, i.e. if X is a Lévy process of the Augmented CGMY class under $\mathbb{P}$, a model class preserving measure would cause X to still be an Augmented CGMY process under $\mathbb{Q}$ (possibly with different parameters). The mean correcting martingale measure (as described in the second chapter) just changes the drift of the Lévy process to move from $\mathbb{P}$ to $\mathbb{Q}$. Unfortunately this does not preserve our model in general.

The Esscher transform however can preserve our model. When we define our model under $\mathbb{P}$ by its Lévy density $\nu_{\mathbb{P}}(x) = C|x|^{-1-Y}e^{-G(x)|x|}$, then under $\mathbb{Q}$ we define the Lévy density by $\nu_{\mathbb{Q}}(x) := e^{\theta x}\nu(x)$. Expanding this, we have

$$\nu_{\mathbb{Q}}(x) := \begin{cases} C|x|^{-1-Y}e^{-(G(x)+\theta)|x|} & \text{if } x < 0 \\ C|x|^{-1-Y}e^{-(G(x)-\theta)|x|} & \text{if } x > 0 \end{cases}.$$

We can immediately see that this is a an Augmented CGMY model with $C_{\mathbb{Q}} = C_{\mathbb{P}}$, $Y_{\mathbb{Q}} = Y_{\mathbb{P}}$ and

$$G_{\mathbb{Q}}(x) = \begin{cases} G_{\mathbb{P}}(x) + \theta & \text{if } x < 0 \\ G_{\mathbb{P}}(x) - \theta & \text{if } x > 0 \end{cases}.$$

So if we choose θ to be adhering to the constraint $-\min_{x<0} G_{\mathbb{P}}(x) \le \theta \le \min_{x>0} G_{\mathbb{P}}(x)$, we are sure for X to be a Lévy process under $\mathbb{Q}$.

6.6 Calibration example

Considering the DAX Jan-2012 Call option on 08-Dec-2011 at 11:00 with about 8 days to maturity left, we fit the CGMY and the Augmented CGMY model (with $\theta_r = \theta_l = 0.1$ fixed). This shows (compared to the original CGMY model) the following results.

DAX Jan-2012 Call options on 08-Dec-2011 at 11:00 CET calibrated by CGMY and Augmented CGMY

Figure 6.1: CGMY - distance to mid

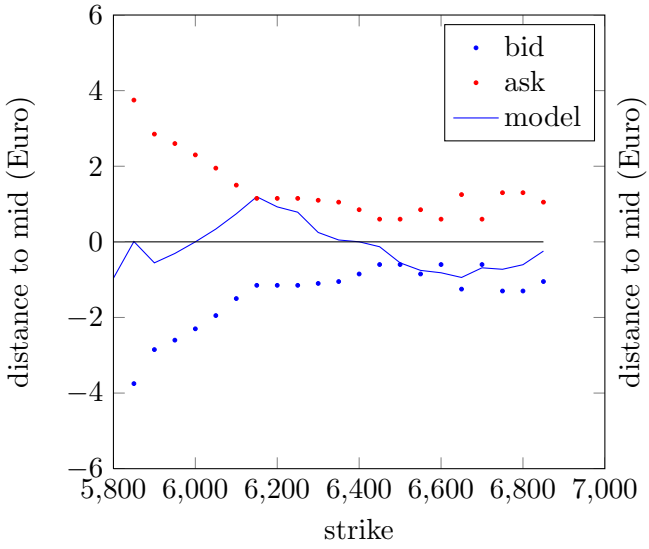


Figure 6.2: Augmented CGMY - distance to mid

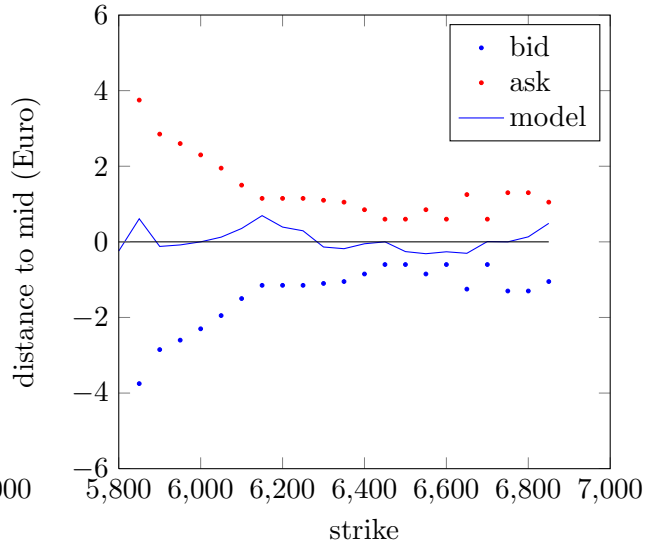


Figure 6.3: CGMY - implied volatility

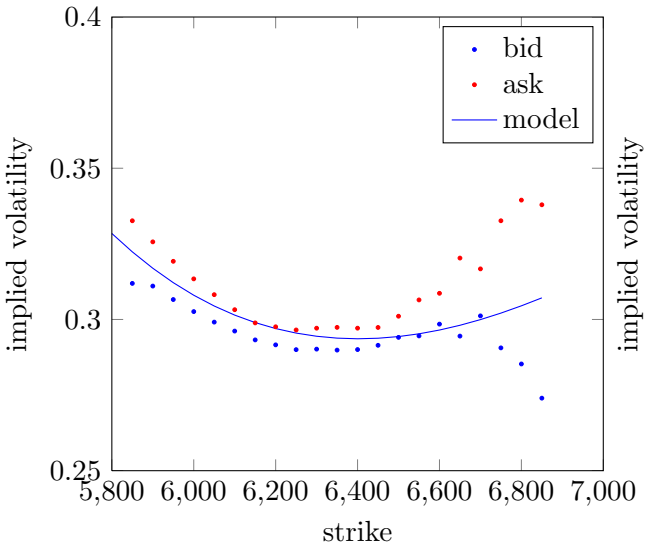
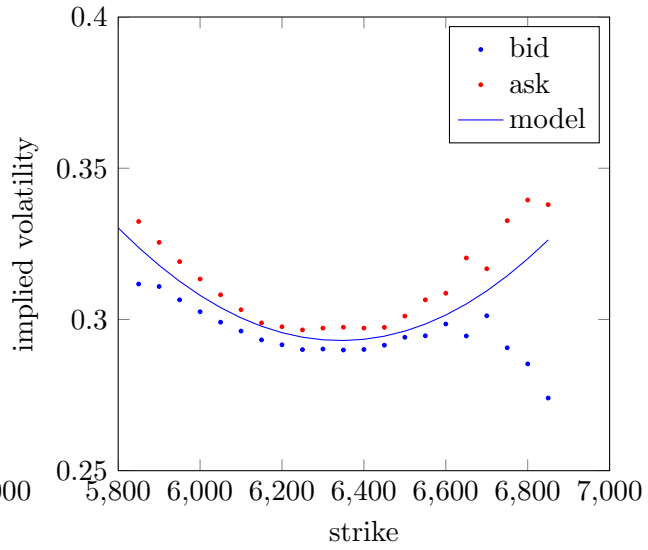


Figure 6.4: Augmented CGMY - implied volatility



CGMY calibrates to the parameters $C = 0.09, G = 4.68, M = 23.04, Y = 1.50$, while Augmented CGMY calibrates to the parameters $C = 0.49, G = 13.60, M = 34.89, Y = 1.18, L = 7.03, R = 6.40$. The Augmented CGMY model clearly shows an improvement for this particular slice of data. The tails of both put and call options fit better, and the fit is entirely between the bid-ask spread. Mind that the maturity of this option is quite short; the Augmented CGMY model tends to outperform the CGMY model especially for the short maturities (as the implied volatility curve becomes more curvy and the Augmented CGMY model has more parameters to compensate for that).

6.6.1 Stability of the parameters

A desirable property for a model is having stable parameters. To investigate this, we fit the DAX Jan-2012 options with a maturity of 16 days left during the day of 04-Jan-2012 (the DAX index was trading between 6088.06 and 6163.48), monitored every minute between 9:00 and 17:30 CET.

On the next two pages we see the futures price (indicative of the price of the index), the loss function (Absolute distance to mid), the percentage of observations that are within the spread, and the fitted parameters of the Augmented CGMY model with 8 parameters, as they change throughout the day. The first thing to notice is that the model fits our data quite well in the sense of being within the spread - we see that generally over 94% of the observations are within the bid-ask spread. There are two occasions where the absolute distance to mid becomes higher (around 0.5 and 0.7 compared to 0.2-0.3 elsewhere). As at those times the observations are still 98% and 100% within the spread, we conclude that at those times the bid-ask spread was wider than usual. We usually see this in the market when an anticipated news event is about to be released.

What we see furthermore is that the parameters of the model behave quite volatile. The meaning of this could be twofold. First, we would expect the moments of the probability distribution of the index at maturity, as implied by the traded option prices, to change in a relatively smooth way over time. Therefore if the parameters change in a volatile way, there could be a chaotic relationship between the two. A second reason could be the used loss function. As there are multiple possible ways to fit one slice of data, our particular choice through time could cost us some parameter stability.

In the table associated with these graphs, we see a comparison of the fits of different versions of the Augmented CGMY model. The Mean Distance to the Mid (MDM), the Mean Average Distance to the Mid (MADM) and the Percentage of Observations within the Spread (%OWS) are compared to the number of parameters of the model. We see that adding two parameters to the CGMY model improves the fit from 0.35 euros from the mid (MADM) to 0.26 euros for the 6 parameter model. However adding another two parameters barely improves this result. As the calibration time increases significantly with respect to the number of parameters, we could see the 6 parameter model as the most useful from these four models.

DAX options during 04-Jan-2012 at every minute between 9:00 and 17:30 CET

Figure 6.5: DAX Mar-2012 Future

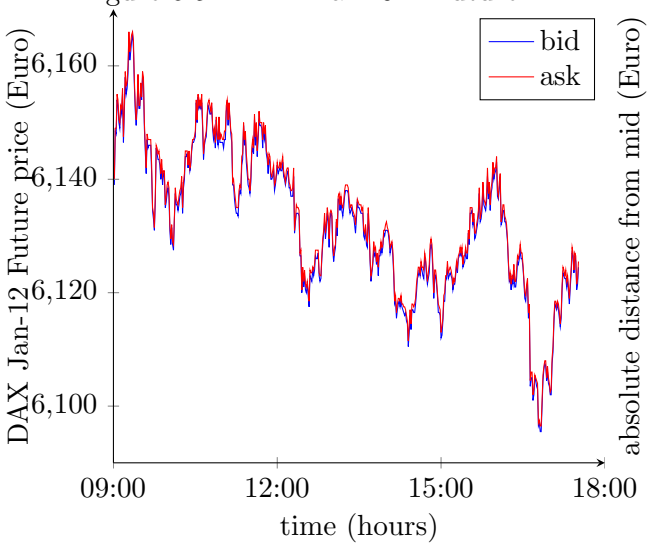


Figure 6.6: Absolute distance to mid

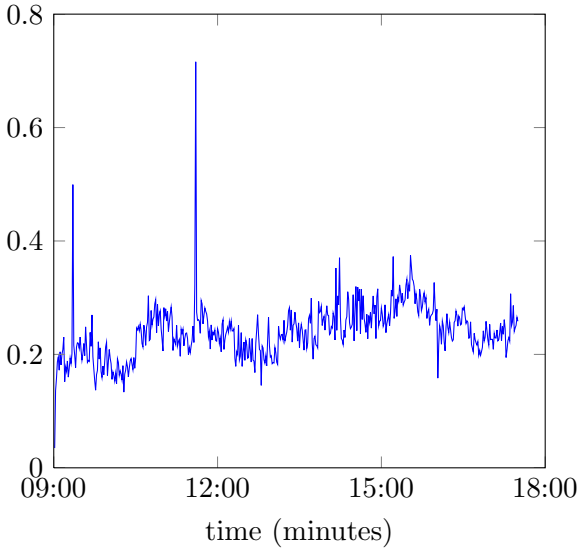


Figure 6.7: Percentage of observations within spread

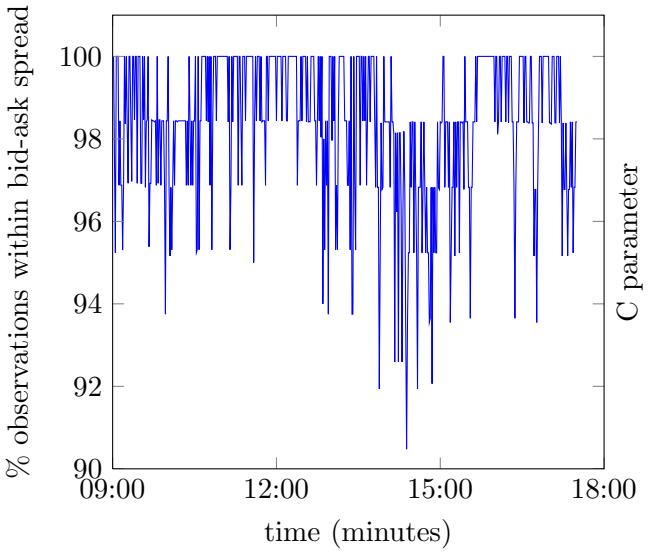


Figure 6.8: C parameter

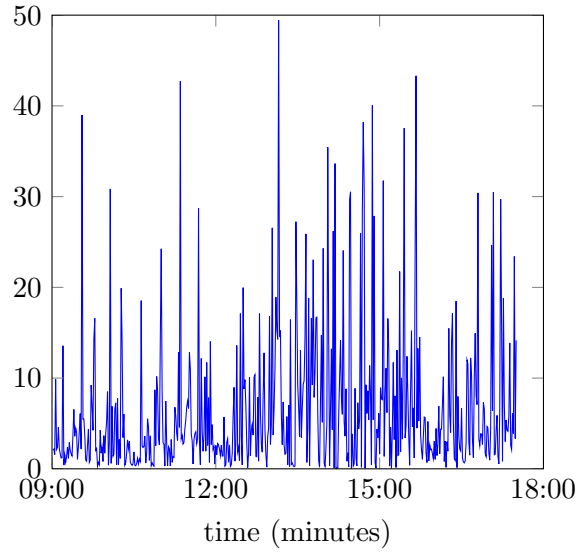


Figure 6.9: G parameter

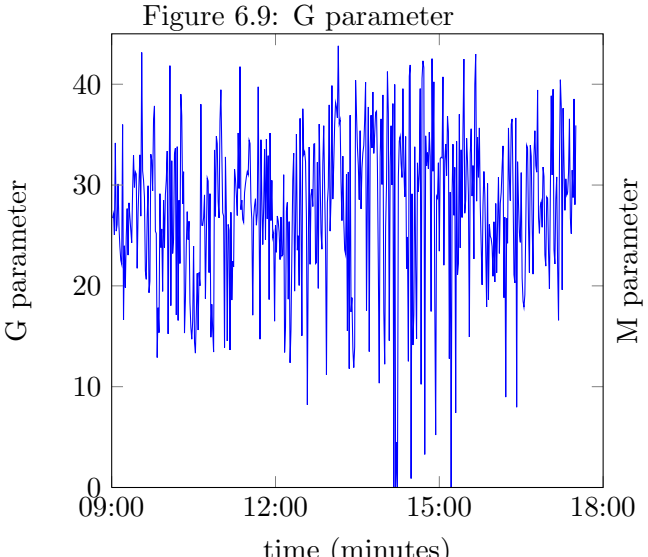
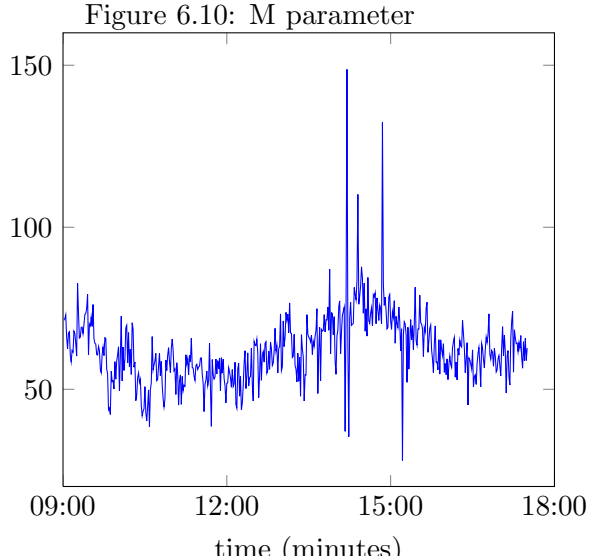


Figure 6.10: M parameter



DAX options during 04-Jan-2012 at every minute between 9:00 and 17:30 CET

Figure 6.11: Y parameter

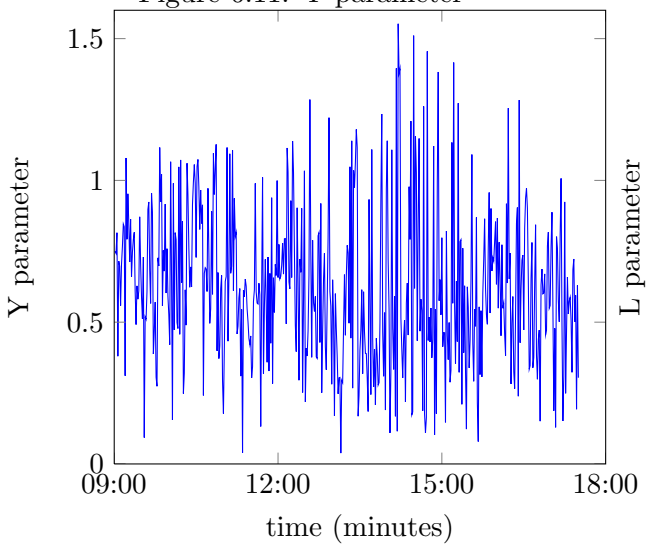


Figure 6.12: L parameter

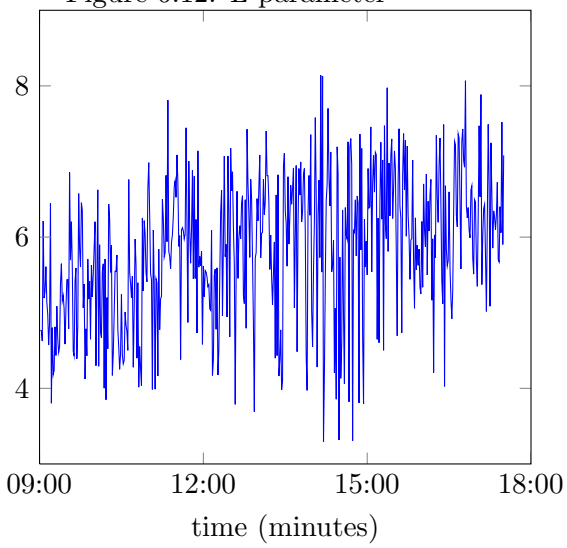


Figure 6.13: R parameter

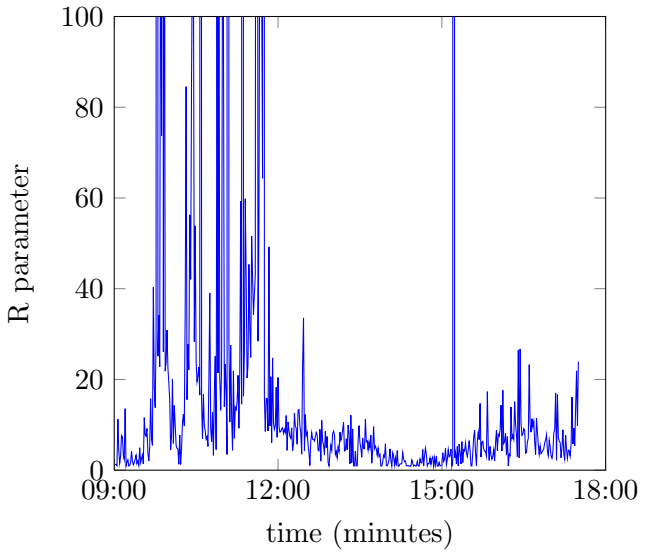


Table of results

Model	MDM	MADM	%OWS	# parameters
CGMY	0.10	0.35	92.9%	4
Augmented CGMY (5)	0.09	0.28	94.3%	5
Augmented CGMY (6)	0.07	0.26	95.8%	6
Augmented CGMY (8)	0.06	0.26	96.3%	8

Results displayed per delta

Figure 6.14: Augmented CGMY: Delta vs. absolute distance from mid per model

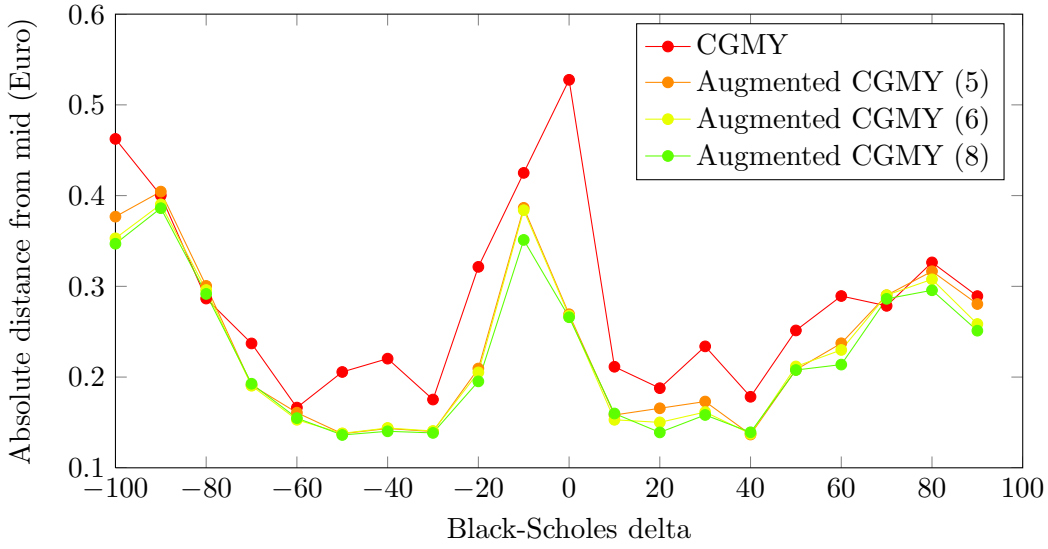
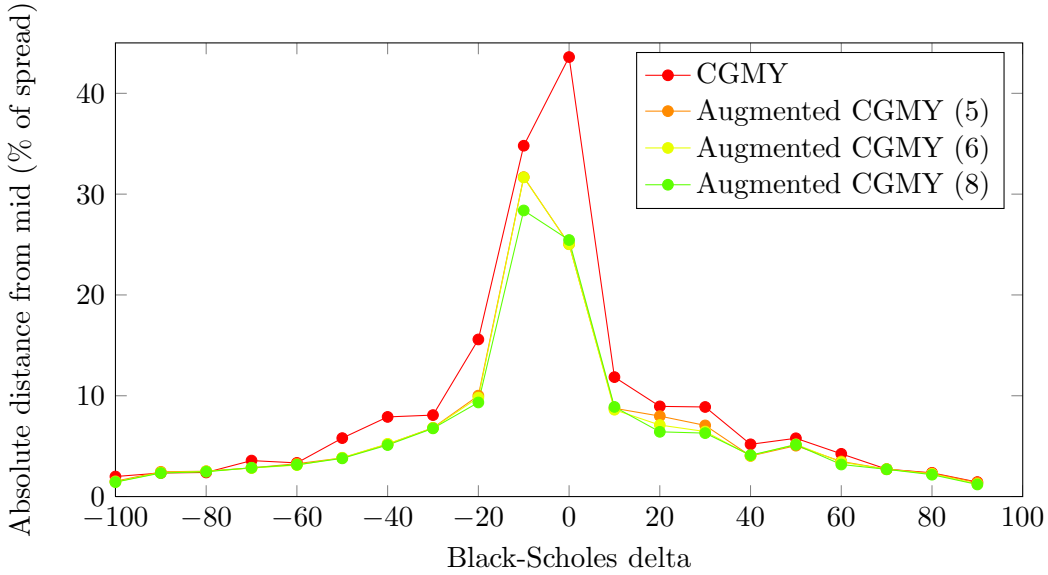


Figure 6.15: Augmented CGMY: Delta vs. absolute distance from mid per model



6.7 Conclusion

We have stated a definition for a new model, based on the CGMY model, with adjusted tails. This model allows for a fit of DAX index options between the bid ask spread, for short maturities. We have stated a few methods to integrate the Lévy density to the characteristic function, and thus be able to use the model to price options (using the COS method). The Double Exponential method is our method of choice in this.

In the table of results we see that the added parameters of the Augmented CGMY model give a better fit. Figures (6.6.1) and (6.6.1) show the calibration results in terms of Black Scholes

delta. The second graph shows the average distance to the mid as a percentage of the spread in Euros. It shows that the region where the models fit the least, at the money (the deltas between -20 and 20), is improved significantly by the Augmented CGMY model.

The parameters of the Augmented CGMY model are not stable at all, as seen in the parameter plots measured over one day. There can be many causes for this, but the most likely is that that the correspondence between the moments of the Augmented CGMY process and its moments is quite chaotic (the moments should be stable over time). This is not necessarily a problem, but if we were to use the model in practice, we should perhaps try to find a quick correspondence between the moments and the parameters - or perhaps force the parameters to stay stable over time.

Another thing that catches our attention is the small effect of the two added parameters in the 8-parameter Augmented CGMY model versus the 6-parameter version. This could show a limitation of our ability to calibrate a model with so many parameters. As the dimensionality increases it becomes of course harder to find an optimum in a reasonable amount of time. Therefore we would best stick to our 6 parameter model.

Bibliography

M. Abramowitz and I.A. Stegun. Handbook of mathematical functions. *Washington, DC: National Bureau of Standards*, 1965.

O.E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and stochastics*, 2 (1):41–68, 1997.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *The journal of political economy*, pages 637–654, 1973.

P. Carr and D. Madan. Option valuation using the Fast Fourier transform. *Journal of Computational Finance*, 2(4):61–73, 1999.

P. Carr, H. Geman, D.B. Madan, and M. Yor. The fine structure of asset returns: an empirical investigation. *The Journal of Business*, 75(2):305–332, 2002.

R. Cont and P. Tankov. *Financial modelling with jump processes*, volume 2. Chapman & Hall, 2004.

R. Cont, P. Tankov, and E. Voltchkova. Hedging with options in models with jumps. *Stochastic analysis and applications*, pages 197–217, 2007.

F. Fang and K. Oosterlee. A novel pricing method for European options based on Fourier-cosine series expansions. 2008.

W. Gautschi. Construction of Gauss-Christoffel Quadrature Formulas. 1968.

G.H. Golub and J.H. Welsch. Calculation of Gauss quadrature rules. 1967.

J.M. Harrison and S.R. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11(3):215–260, 1981.

E. Lukacs. *Characteristic functions*. Griffin, 1960.

D.B. Madan, P.P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998.

J.R.R.A. Martins, P. Sturdza, and J.J. Alonso. The complex-step derivative approximation. *ACM Transactions on Mathematical Software (TOMS)*, 29(3):245–262, 2003.

K.I.M. McKinnon. Convergence of the Nelder-Mead simplex method to a nonstationary point. *SIAM Journal of Optimization*, 9:148–158, 1998.

R.C. Merton. Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, pages 141–183, 1973.

I. Monroe. Processes that can be embedded in Brownian motion. *The Annals of Probability*, pages 42–56, 1978.

M. Mori. Discovery of the double exponential transformation and its developments. *Publications of RIMS, Kyoto University*, 41(4):897–935, 2005.

J. Nocedal and S.J. Wright. *Numerical optimization*. Springer verlag, 1999.

T. Ooura. A double exponential formula for the Fourier transforms. *Publ. Res. Inst. Math. Sci*, 41(4):971–977, 2005.

P.E. Protter. *Stochastic integration and differential equations*. Springer Verlag, 2004.

K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

W. Schoutens. *Lévy processes in finance: pricing financial derivatives*. Wiley, 2003.

W. Schoutens and J.L. Teugels. Lévy processes, polynomials and martingales. *Stochastic Models*, 14(1-2):335–349, 1998.

S.E. Shreve. *Stochastic Calculus for Finance II: Continuous Time Models*. Springer Finance, 2005.

J.M. Steele. *Stochastic calculus and financial applications*, volume 45. Springer Verlag, 2001.

M. Sugihara. On the optimality of the DE transformation. *Kokyuroku RIMS, Kyoto University*, (585):150–175, 1986.

H. Takahasi and M. Mori. Double exponential formulas for numerical integration. *Publications of the Research Institute for Mathematical Sciences*, 9(3):721–741, 1974.

P. Tankov. Lévy processes in finance and risk management. *Wilmott Magazine, September-October*, 2007.

P. Tankov. Pricing and hedging in exponential Lévy models: review of recent results. *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 319–359, 2011.

L. Yao, G. Yang, and X. Yang. A note on the mean correcting martingale measure for geometric Lévy processes. *Applied Mathematics Letters*, 24(5):593 – 597, 2011.