Fitting Financial Models to Market Data Using Kriging

Eelse-jan Stutvoet

November 11, 2007

Summary

In this thesis we attempt to fit financial models to market data consisting of the bid and ask prices for American call and put options for various strikes. As these are dividend paying assets we aim at determining the height of this payment from the market data.

To model this assets with dividend we will use the Korn-Rogers approach, which models the asset price based on these dividend payments. We use three models to the dynamics of these dividend payments, the Black-Scholes model, the Variance Gamma model and the CGMY model.

To calculate the option prices from this asset price model we use the CONV method, a Fast Fourier Transform method. To determine how well a parameter set fits the market data we use a weighted squared error function.

We use the Kriging method to search for a parameter set for which the error function obtains the minimum value. The Kriging method models the surface of the error function based on the evaluated points so far. For each point in the surface it returns a normal distribution. Based on these distributions we introduce a stopping criterion. As the Kriging model gets quadratically slower with the number of evaluated points added, we need to break off the algorithm at some point. In order to still get good results, we need a search range which is small and smooth enough to find the optimal solution before the algorithm breaks off.

The Black-Scholes driven model can not be calibrated to fit the market data well enough. To calibrate the Variance Gamma model we need to use a two-step method where we zoom in on the interesting region to get stable results for the optimal value. This is time-expensive. The CGMY driven model needs the Variance Gamma best fit to determine an initial search space or the model or the algorithm breaks off too soon to give us an interesting region to search. Even using this method, we still can not get completely stable results even when we zoom in twice. This leaves the Variance Gamma driven model as the only workable model we have considered.

The optimal parameter sets for the Variance Gamma driven and the CGMY driven model stay valid for about two weeks on the data sets we examined.

We can not obtain very accurate implied dividends from a single data set. The results differ by € 0.06 for parameter sets which all stay within the bid-ask spread.

Contents

1   Introduction

Many models exist to calculate option prices. One problem common to all models is fitting the model to reality by determining an optimal parameter set. Of course it is hard to determine which parameter set will best describe the behavior of the asset and option prices in the future. One way to determine such a parameter set, in fact the way we will look at it in this thesis, is to determine the parameter set which best fits market data.

Market data consists of the bid and ask spreads market makers have to supply for each option. This does not necessarily mean that this is the only correct price for the option, but the market mechanism is supposed to correct prices which are either too high or too low. Market data reflects the price the market makers think the option is worth. In particular, we can determine from this market data the height of the dividend market makers suspect the asset to have.

There are black-box programs available which in some way extract a parameter set from the market data. In this thesis we want to investigate some means to determine such parameter sets ourselves, and try and get some insight in the problem.

To this end we will implement the Kriging algorithm [1] which is basically a general way to determine an optimum for any deterministic function. Kriging works by taking a number of points evenly spread over the search domain, evaluating the value of the function in those points and approximating the surface of the function using these values. This approximation surface is searched for new points to evaluate. These new points are incorporated in the approximation which becomes a better approximation of the real function the more points are added. At some point the evaluated point with the optimal value is presented as the solution.

The market data consists of the bid and ask prices for simple American call and put options. A put option is a contract that gives the owner the privilege to sell a share of stock at a predetermined price, the so-called strike price. A call option permits the owner to buy a share of stock at a predetermined price. In the option contract a date is named after which the contract is void, the expiry date. There are some differences in option contracts based on when the option can be exercised. For example, European options can only be exercised at the expiry date. American options can be exercised at any time before expiry.

One well-known model to price options on assets without dividend payments is the Black-Scholes model [2]. The model has only one parameter (we assume the interest rate r to be known), the volatility, which completely determines the relation between the asset price and the option price (given an expiry date and a strike price). Unfortunately, this is not realistic. When determining the implied volatilities from options with the same expiry date and different strike prices one would expect to get a straight line if the model

was correct. In practice this graph would show a skew, the so-called volatility smile. One difference between model and reality which could explain this behavior is the number of jumps of several standard deviations which occur in the market. This behavior cannot be modeled by the Normal distribution which drives the Black-Scholes model. The tails of the Normal distribution vanish exponentially, which does not correspond to reality. Also, the Black-Scholes model does not allow for jumps at all, while these jumps can be observed in markets.

There are several models which address this problem in the Black-Scholes model. In this thesis we will look at two Lévy jump models, Variance Gamma [3] and CGMY [4]. In these models the asset price can jump and the tails of the distribution are fatter than those of the Normal distribution. Using these models it should be possible to model this volatily smile and fit the model to a range of strike prices at once.

Unfortunately the Black-Scholes, Variance Gamma and CGMY models do not deal with dividend payments. In order to incorporate dividend payments we will use the Korn-Rogers approach [5]. In this approach, the asset price is not modeled directly, but as the sum of the expected dividend payments. We can use the above models as dynamics for these dividend payments.

Even given a model for the asset price, there are still several methods to calculate option prices from these models. We will calculate the option prices using the CONV method [6], a Fast Fourier Transform model.

In this thesis we will evaluate the effectiveness of the Kriging approach on the three models. We will examine how well the models can fit the market data and if the optimal parameter sets we find are stable in time. In these parameter sets we will put some extra focus on how well they can predict the implied dividend.

2 Notation

Throughout this document $\boldsymbol{\theta}$ denotes a set of k parameters of the model we try to fit to the market data.

$$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \tag{1}$$

The last parameter will always denote the dividend rate $\theta_k = \delta$.

Each set represents a point in our search space. In the Kriging model we consider n such sets and will distinguish them by writing $\boldsymbol{\theta}^{(i)}$ to indicate the ith set.

In the financial models we consider, the asset price is denoted by the random variable S, or S_t if a more explicit time distinction is required. When we use the risk-neutral asset price, we will explicitly indicate this by S_{RN}.

7

In the calculations we will use the log-transformed risk-neutral asset price $X = \ln S_{\mathrm{RN}}$. Again we will sometimes add a subscript to indicate the time dependence more explicitly. The financial models are distinguished by the characteristic function $\phi(u, t, \boldsymbol{\theta})$ of the density of X_t. When discussing the derivation of this function for a fixed parameter set we will drop $\boldsymbol{\theta}$ from this notation. The specific model will be indicated by a subscript where this is relevant: $\phi_{\mathrm{BS}}(u, t)$.

In the assessment phase we will use an objective function $g(\boldsymbol{\theta})$ to assign a value to each set. Although g is written as a function of only $\boldsymbol{\theta}$, it is in fact dependent on $\phi(u, t)$ and the market data as well:

$$g(\boldsymbol{\theta}) = g(\phi(u, t, \boldsymbol{\theta}), \mathrm{market\ data}) \tag{2}$$

This value $g(\boldsymbol{\theta})$ can be considered as a measure of the distance of the model using the parameter set $\boldsymbol{\theta}$. We will call this the error of this solution. Because the market data contain uncertainty in the form of a gap between bid and ask prices, we denote this error on the inputs as ϵ_0 and provide it as a reference whenever results are presented.

The minimum value of the set of sample points will be denoted as $g_{\min}$.

$$g_{\min} = \min_{i \in \{1, \ldots, n\}} \{g(\boldsymbol{\theta}^{(i)})\} \tag{3}$$

3 The Kriging model

In this chapter we will describe the Kriging model. We will start to discuss the whole framework. In the following sections we will examine the sampler and the genetic algorithm in some more detail.

3.1 Kriging

Kriging is a method to model the surface of a deterministic function with k parameters. This surface is constructed by sampling at n points. It mimics the optimization landscape, represented by an objective function. The surface represents a surrogate for the true objective function. It is continuously updated by including more points to its representation.

   The Kriging model provides not only an estimate of the value at any specific point but also an indication of the error. We use an optimization algorithm that takes both of these values into account to find additional sample points to improve our estimate of the optimal value as described in [9].

   In this section we will discuss the origin of the model, show how it works and propose a stopping criterion. At the end of this section we will show the algorithm in a flowchart.

3.1.1 Origin of the Kriging model

A well-known method to fit a response surface to n function evaluations is linear regression. Let $\boldsymbol{\theta}^{(i)} = (\theta_1^{(i)}, \ldots, \theta_k^{(i)})$ denote the ith sample point and $g^{(i)} = g(\boldsymbol{\theta}^{(i)})$ the corresponding value of the evaluation function. The samples are now treated as realizations from a stochastic model:

$$g(\boldsymbol{\theta}^{(i)}) = \sum_{h=1}^{k} \beta_h l_h(\boldsymbol{\theta}^{(i)}) + \epsilon^{(i)}$$

   In this model, each $l_h(\boldsymbol{\theta})$ represents a linear or nonlinear continuous function of $\boldsymbol{\theta}$, the β_h's are unknown coefficients to be estimated and the $\epsilon^{(i)}$ are normally distributed, independent error terms with mean zero.

This approach may be appropriate for physical measurements, but when dealing with a deterministic function, as in our problem, there are some modifications necessary.

1. The error terms can not be independent of $\boldsymbol{\theta}$.

   Since the function $g(\boldsymbol{\theta})$ is deterministic, any error term will not be random, but rather a leftover term in $\boldsymbol{\theta}$. From now on, we will write $\epsilon(\boldsymbol{\theta}^{(i)})$ instead of $\epsilon^{(i)}$.

2. The error terms can not be independent of each other.

   If $g(\boldsymbol{\theta})$ is a continuous function, this means that $\epsilon(\boldsymbol{\theta})$ should also be continuous as it is the difference between $g(\boldsymbol{\theta})$ and the continuous regression terms. This implies that $\epsilon(\boldsymbol{\theta}^{(i)})$ and $\epsilon(\boldsymbol{\theta}^{(j)})$ are highly correlated when $x^{(i)}$ and $x^{(j)}$ are "close".

These reservations lead to a different model. Instead of estimating the weights of the regression terms, we will estimate weights in a correlation structure. The regression terms are replaced by a constant term.

The sample points are now considered as realizations of the following stochastic model:

$$g(\boldsymbol{\theta}^{(i)}) = \mu + \epsilon(\boldsymbol{\theta}^{(i)}) \qquad (i = 1, \ldots, n), \tag{4}$$

where μ is the mean of the stochastic process, $\epsilon(\boldsymbol{\theta}^{(i)})$ is Normal$(0, \sigma^2)$ and correlations between processes are given by

$$\mathrm{Corr}[\epsilon(\boldsymbol{\theta}^{(i)}), \epsilon(\boldsymbol{\theta}^{(j)})] = \exp[-d(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(j)})], \tag{5}$$

where the "distance" is defined as

$$d(\boldsymbol{\theta}^{(i)}, \boldsymbol{\theta}^{(j)}) = \sum_{h=1}^{k} w_h |\theta_h^{(i)} - \theta_h^{(j)}|^2 \qquad (w_h \geq 0). \tag{6}$$

The w_hs represent weighting factors that are added to each parameter.

3.1.2 Constructing the initial Kriging model

In our final model we have $k + 2$ parameters:

$$\mathbf{w} = (w_1, \ldots, w_k), \qquad \mu, \qquad \sigma^2 \tag{7}$$

To estimate these parameters we maximize the likelihood of our sample points. In order to obtain a compact expression we introduce the following notation.

We define $\mathbf{g}$ to be the n-vector of the objective values of all n sample points:

$$\mathbf{g} = (g(\boldsymbol{\theta}^{(1)}), \ldots, g(\boldsymbol{\theta}^{(n)})). \tag{8}$$

$\mathbf{R}$ denotes the $n \times n$-correlation matrix:

$$\mathbf{R}_{i,j} = \mathrm{Corr}[\epsilon(\boldsymbol{\theta}^{(i)}), \epsilon(\boldsymbol{\theta}^{(j)})]. \tag{9}$$

We write $\mathbf{1}$ for an n-vector of ones.

Using this notation we obtain the following expression for the likelihood function:

$$\ell(\mathbf{w}, \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}|\mathbf{R}|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{g} - \mathbf{1}\mu)'\mathbf{R}^{-1}(\mathbf{g} - \mathbf{1}\mu)}{2\sigma^2}\right] \quad (10)$$

For fixed values of $\mathbf{w}$ we can calculate the values of μ and σ^2 that maximize the likelihood function:

$$\hat{\mu} = \frac{\mathbf{1}'\mathbf{R}^{-1}\mathbf{g}}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}} \quad (11)$$

and

$$\hat{\sigma}^2 = \frac{(\mathbf{g} - \mathbf{1}\hat{\mu})\mathbf{R}^{-1}(\mathbf{g} - \mathbf{1}\hat{\mu})}{n} \quad (12)$$

Substituting (11) and (12) in (10) we obtain a k-dimensional unconstrained nonlinear optimization problem.

We solve this problem by using an iterative method as suggested in [7]

$$\mathbf{w}^{(i+1)} = \mathbf{w}^{(i)} + B^{-1}\frac{d\ln\ell}{d\mathbf{w}} \quad (13)$$

where

$$B_{ij} = \frac{1}{2}\mathrm{trace}\left(\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial w_i}\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial w_j}\right) \quad (14)$$

and

$$\frac{\partial\ln\ell}{\partial w_i} = -\frac{1}{2}\mathrm{trace}\left\{\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial w_i}\right\} + \frac{(\mathbf{g} - \mathbf{1}\hat{\mu})\mathbf{R}^{-1}\frac{\partial\mathbf{R}}{\partial w_i}\mathbf{R}^{-1}(\mathbf{g} - \mathbf{1}\hat{\mu})}{2\hat{\sigma}} \quad (15)$$

If (13) is not an improvement, we use an updated version of B

$$B^{\mathrm{up}} = B + 2^d\mathrm{diag}B \quad (16)$$

for increasing numbers of d as described in [8]. This will ensure that each step is an improvement. If the improvements get very small (¡0.1%) we stop the search and consider the algorithm converged.

Once we have found the optimal weights, we can estimate the value of the objective function for arbitrary $\boldsymbol{\theta}^*$ using the following linear unbiased predictor:

$$\hat{g}(\boldsymbol{\theta}^*) = \hat{\mu} + \mathbf{r}'\mathbf{R}^{-1}(\mathbf{g} - \mathbf{1}\hat{\mu}) \quad (17)$$

where $r_i(\boldsymbol{\theta}^*) \equiv \mathrm{Corr}[\epsilon(\boldsymbol{\theta}^*), \epsilon(\boldsymbol{\theta}^{(i)})]$ is calculated using (6) and (5).

Likewise we can obtain an estimate of the accuracy of our solution by the formula for the mean squared error of the predictor, denoted by $s^2(\boldsymbol{\theta}^*)$:

$$s^2(\boldsymbol{\theta}^*) = \hat{\sigma}^2\left[1 - \mathbf{r}'\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}'\mathbf{R}^{-1}\mathbf{r})^2}{\mathbf{1}'\mathbf{R}^{-1}\mathbf{1}}\right] \quad (18)$$

3.1.3 Finding the global optimum

Once the initial Kriging model is constructed we use a genetic algorithm to determine a next optimal point to evaluate.

Since the model includes a measure of uncertainty (18) in a predicted point, it would be possible to oversee the global optimum if we only took the predicted value (17) into consideration. An improved way to search for the optimum is to take the uncertainty into account. This leads to the so-called search criterion of the Expected Improvement (EI) [9].

To calculate the EI, we consider $g(\boldsymbol{\theta}^*)$ to be a random variable with normal distribution, mean $\hat{g} = \hat{g}(\boldsymbol{\theta}^*)$ and variance $s^2 = s^2(\boldsymbol{\theta}^*)$. Under these assumptions the expectation of the improvement on the current minimum value $g_{\min}$ of our sample set is

$$
\begin{aligned}
\mathbb{E}[I(\boldsymbol{\theta}^*)] &= \mathbb{E}[\max(g_{\min} - g(\boldsymbol{\theta}^*), 0)] \\
&= \int_{-\infty}^{\infty} (g_{\min} - x) 1_{\{x < g_{\min}\}}(x) \frac{1}{s} f(\frac{x - \hat{g}}{s}) dx \\
&= \int_{-\infty}^{g_{\min}} (x - g_{\min}) \frac{1}{s} f(\frac{x - \hat{g}}{s}) dx \\
&= \int_{-\infty}^{\frac{g_{\min} - \hat{g}}{s}} (g_{\min} - sz - \hat{g}) f(z) dz \\
&= (g_{\min} - \hat{g}) \int_{-\infty}^{\frac{g_{\min} - \hat{g}}{s}} f(z) dz - s \int_{-\infty}^{\frac{g_{\min} - \hat{g}}{s}} z f(z) dz
\end{aligned}
\tag{19}
$$

where $f(x)$ is the density function of the standard normal distribution. Since

$$
\begin{aligned}
\int_{-\infty}^{a} z f(z) dz &= \int_{-\infty}^{a} z \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz \\
&= [-\frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}]_{-\infty}^{a} \\
&= -f(a)
\end{aligned}
\tag{20}
$$

we can write the following expression for the EI:

$$
\mathbb{E}[I(\boldsymbol{\theta}^*)] = (g_{\min} - \hat{g}) F((g_{\min} - \hat{g})/s) + s f((g_{\min} - \hat{g})/s)
\tag{21}
$$

where F denotes the cumulative distribution function of the standard normal distribution.

We now calculate the true value of the point with the highest EI and add this as an extra sample point to the model. In this way we combine searching for a global optimum with improvement of the approximation surface in relevant areas.

3.1.4 Stopping condition and validation

In general the Kriging method shows a descend in the maximum EI found in each step. Sometimes it goes up when a new interesting region has been found, but once explored, it drops further. Though this is certainly a sign of convergence, when can we be reasonably sure the optimum has been found?

One way of determining how well our approximation surface fits the objective function is to leave out a sample point and compare the prediction of the model without this point to the true value in terms of the predicted standard deviation. If all sample points are predicted within 3 standard deviations of their true value we consider the model to fit well enough.

To formulate the stopping condition we assume that if all initial sample points are predicted within 3 standard deviations of their true value, this property will hold for all the predicted points in the domain $\mathcal{D}$. In formula form:

$$|g - \hat{g}| < 3s \qquad \forall g \in \mathcal{D} \tag{22}$$

If this assumption holds we can stop the algorithm once

$$g_{\min} - \hat{g} < -3s \qquad \forall g \in \mathcal{D} \tag{23}$$

as all points in our approximation at are more than 3 standard deviations distance of our current minimum.

For the maximum EI it holds that if

$$(-3F(-3) + f(-3))\max_{\mathcal{D}} s \leq \max_{\mathcal{D}} \mathrm{EI} \tag{24}$$

then

$$\frac{g_{\min} - \hat{g}}{s} \geq -3 \tag{25}$$

since

$$\max_{\mathcal{D}} \mathrm{EI} = \max_{\mathcal{D}} (g_{\min} - \hat{g})F((g_{\min} - \hat{g})/s) + sf((g_{\min} - \hat{g})/s) \tag{26}$$

$$\leq \max_{\mathcal{D}} s \max_{\mathcal{D}} \frac{g_{\min} - \hat{g}}{s} F((g_{\min} - \hat{g})/s) + f((g_{\min} - \hat{g})/s)$$

and $xF(x) + f(x)$ is a strictly increasing function.

As an estimate for $\max_{\mathcal{D}} s$ we will use $\hat{\sigma}$ and stop improving by EI once $\max_{\mathcal{D}} \mathrm{EI} < \hat{\sigma}(-3F(-3) + f(-3))$.

Note that when this criterion on EI is reached, (23) is not necessarily fulfilled. However, if it does not hold then (23) has certainly not been reached.

To make sure that (23) holds, we can add points with the highest number of standard deviation distance $\frac{g_{\min} - \hat{g}}{s}$ to the model. Once the highest standard deviation distance is smaller than -3 we have reached our stopping criterion (23). We will call this extra stopping criterion Maximum Variance Distance (MVD). In chapter 5 we will determine if adding this criterion is worth the extra trouble.

3.1.5   Overview

In this section we will give an overview of the total method for finding an optimal parameter set. In the following sections we will go into each of the steps in more detail.

1. We construct an initial sample set.

2. We assess the performance of these sample points by an objective function which basically measures the error of using the option price model with these parameters compared to market data.

3. The performance of the sample points is used to create the initial Kriging model. The model provides an approximation of the performance of any parameter set and gives us an indication of the modeling error.

4. Since the model is an approximation, finding the optimal point in the Kriging model could give us a local optimum or a point that is not an optimum at all. Instead of looking for the point with the lowest predicted value, we try to find the point that is most likely to improve our current best sample point. This new objective function is known as the Expected Improvement (EI). By adding the point with the highest EI to the model, we improve the accuracy of optima and at the same time ensure that points with a great chance of being global optima are not left out.

   A Genetic Algorithm (GA) is used to find the global optimum of the EI.

5. The Kriging model is continuously updated by including new points in the approximate model until we have reached our stopping criterion.

The complete process is depicted in Figure 1. More details on the subsequent steps in the process will be provided in the following sections.

Figure 1: Flowchart for determining the optimal parameter set.

3.2 Sampler

To start the Kriging algorithm, we need some way to select the first sample points. One common way to select sample points is Latin Hypercube Sampling (LHS) [10].

To generate n sample points in this algorithm, we split each axis of our search space in n equal parts and take a sample in each. By combining these results in a random order, we end up with n sample points that are more or less evenly divided over the sample space.

However, this does lead to some clustering in the sample space as can be seen in Figure 2. To get a more even spread of the sample points we can also use the existing Kriging structure to generate sample points. We initiate the algorithm with neutral weights and sample points with the highest uncertainty, thus ensuring a more or less even spread over our sample space. We will call this method Kriging Sampling (KS).

The precise algorithm is present point-wise below.

- As an initial guess the weights w_i are set to $\frac{100}{\sqrt{d}l_i^2}$, where d is the number of dimensions and l_i is the length of the interval for the ith variable.

- To fill the various matrices and vectors, the first points are selected at random.

- Instead of using the Expected Improvement criterion to select new sample points we add the point with the highest variance.

- More and more points are added until the required number of sample points is reached.

While sampling this way is more complex and time consuming than using LHS (it is equally expensive as adding points using the EI criterion), the initial sample set is far superior in spread as can be seen in picture 3. Disadvantages to this method are that we have no guarantee that each axis is evenly sampled and there are relatively many points on the boundary of our domain.

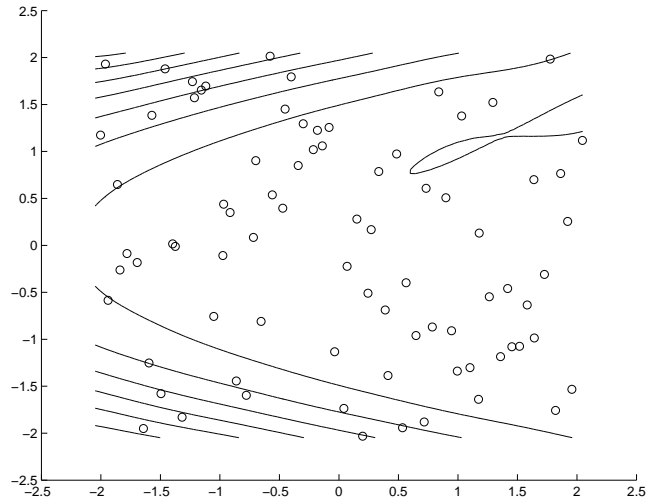In chapter 5 we will compare the two to decide which one to use in the calibration program.

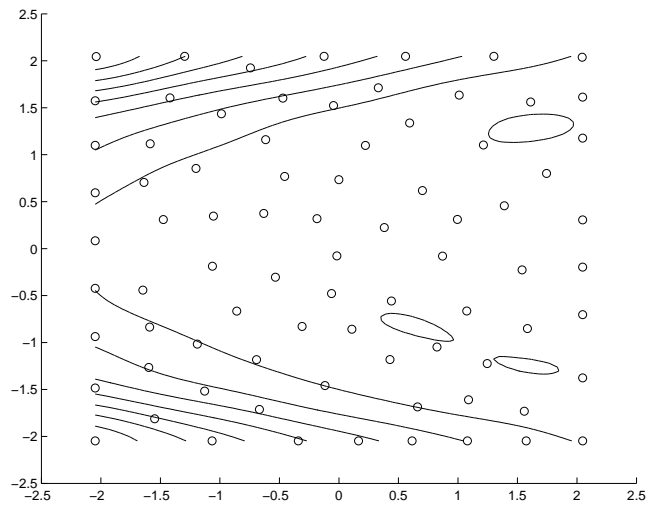Figure 2: Spread of 50 sample points using Latin Hypercube Sampling on the 2-D Rosenbrock function.

Figure 3: Spread of 50 sample points using Kriging sampling on the 2-D Rosenbrock function.

17

3.3 Genetic Algorithm

In the center of the fitting algorithm a Genetic Algorithm (GA) is used to find the point with the greatest expected improvement in the surrogate optimization landscape.

In this section we will discuss the general approach of genetic algorithms and describe our specific implementation.

3.3.1 General approach

GAs mimic genetics to search for the optimal solution. Solutions are represented as strings of bits. To compare how "close" strings are to each other we compare the number of bits in which they differ. Standard binary encoding has the drawback that some numbers that are neighbors differ in more than one byte, for example 3 (011) and 4 (100).

To address this problem we move to a binary-reflected Gray code. This is an encoding of integers with the characteristic that two neighbors differ in only one bit. It is constructed recursively. We start out with 0 and 1. For each interval $[2^i, 2^{i+1}), i \geq 1$ we use all the previous numbers in backward order and add a 1 in front of them.

$$
\begin{array}{c}
000 - 0 \\
001 - 1 \\
\hline
011 - 2 \\
010 - 3 \\
\hline
110 - 4 \\
111 - 5 \\
101 - 6 \\
100 - 7
\end{array}
\tag{27}
$$

and so on. Solutions will now be encoded as

$$
(2, 5) \rightarrow 011111 \tag{28}
$$

The function we'd like to optimize serves as a measure of the fitness of the individuals.

$$
f_{\text{obj}}(x, y) = xy \tag{29}
$$
$$
f_{\text{fitness}}(011111) = f_{\text{obj}}(2, 5) = 10 \tag{30}
$$

To find new solutions, we combine pairs of these strings (parents). Each bit of the new solutions (children) is inherited from one of the two parents (crossover). One way to do this is single-point crossover: a random point somewhere in the string is chosen and all the bits in front of that point

18

are from the first parent and the others from the second. The combination gotten by switching the two parents results in a second child.

$$
\begin{array}{c}
0010|11 \\
1011|01 \\
\hline
0010|01 \\
1011|11
\end{array}
\tag{31}
$$

This reproduction process may be enhanced by adding a chance on mutation, to keep the process from getting stuck in a local optimum.

A new population is now formed by making some combination of parents and children and the process of crossover and mutation starts anew. The process is terminated if some stopping criterion is reached.

3.3.2 Specific implementation

Though all genetic algorithms follow this general structure, there is a lot of freedom in the implementation. We will use a μ-GA [11] with a population size of 5, elitism, tournament parent selection and uniform crossover. The algorithm will restart when the bits of each member of the population differs in fewer than 5% of the bits of the best solution.

The process is depicted in Figure 4. We will describe how the general process is altered.

We start out with only 5 randomly generated strings.

Next, we get to the selection of the parents. We use a process called tournament parent selection. This works in the following manner:

1. The population is ordered in a random way.

2. The fitness of individual 1 is compared with individual 2. The one with the greatest fitness is chosen as parent 1.

3. The fitness of individual 3 is compared with individual 4. The one with the greatest fitness is chosen as parent 2.

4. Crossover is performed with parents 1 and 2.

This process is done twice, to obtain 4 children.

The crossover we use is called uniform crossover. It is performed in the following fashion: for each bit of the child there is 50% chance of inheriting that bit from parent 1 else it inherits from parent 2.

There is no mutation step. Diversity in the solutions is ensured by restarting the process a number of times.
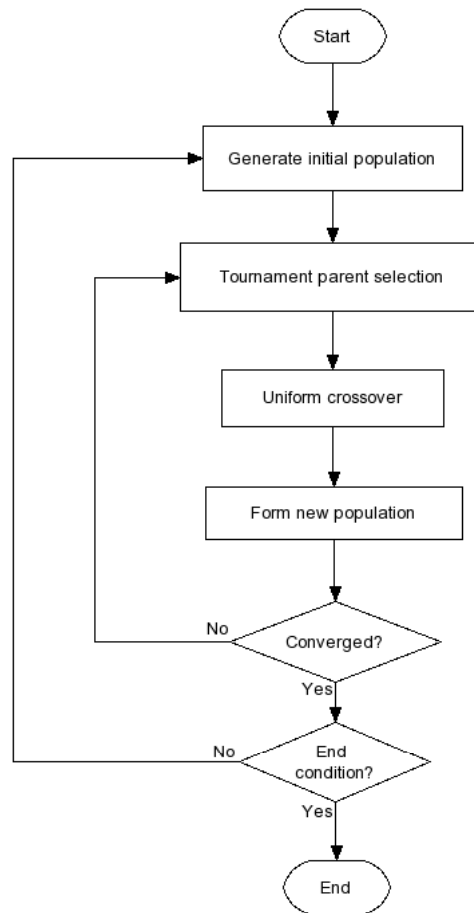
Figure 4: Flowchart of the μ-GA algorithm.

The next generation is formed by the 4 children and the fittest of the parents (elitism).

After a while this process will converge. We consider it converged when the bits of each population member differ in less than 5% from the bits of the fittest member. When this occurs, the algorithm restarts. The fittest member is preserved and 4 new random strings are generated.

When the algorithm converges to the same point 10 times in a row or after 20000 function evaluations which ever occurs first, we stop the algorithm and return the best point so far as the optimal point.

Even if this procedure does not produce the optimal point every time, this will not invalidate our search as those second best points will still be good points to add to the model. In the worst case this will make the search a bit less efficient.

4   Pricing options with dividend payments

In this chapter we will focus on how to price options on stocks with dividend payments.

We will first discuss a general framework to price options without dividend payments which supports various models for the asset price. We follow up this discussion with three models for the asset price. In the final section we will show how the framework can be modified to incorporate dividend payments in the asset and option prices.

4.1   Fast option price evaluation

To raise capital, corporations issue shares of stock. These shares of stock, or stocks for short, are traded on financial markets. To add more flexibility, contracts are made based on future values of these stocks. One example of such a contract is a stock option.

Options are contracts that give the holder the opportunity to buy (a "call" option) or sell (a "put" option) an asset on or before a future date (the "maturity" date) at a predefined price (the "strike" price), all set in the contract. The writer has the obligation to honor the contract. Since the holder gained a right, clearly the option must have some value. As this value is based on the future price of the asset, the model for the option price will be closely related to the model for the development of the asset price.

A commonly used criterion to divide the options into classes is when the option can be exercised. We distinguish the following types of options:

- *The European option.* This option has a single fixed exercise date.

- *The Bermudan option.* In this contact a finite number of exercise dates are fixed.

- *The American option.* This option is the most flexible of the three, as it can be exercised at any time before a fixed end date.

4.1.1   Modeling the asset price

Asset prices are modeled as continuous time random walk processes, so-called Lévy processes. These processes produce sample paths that are cadlag (continue à droite, limite à gauche) which means they have to be continuous from the right and limited from the left. This leads to the following definition:

Definition 1 (Lévy process). *A Lévy process is a cadlag stochastic process* $(X_t)_{t \geq 0}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in* $\mathbb{R}^d$ *such that* $X_0 = 0$ *with the following properties:*

1. *Independent increments: for every increasing sequence of times $t_0, \ldots, t_n$, the random variables $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.*

2. *Stationary increments: the law of $X_{t+h} - X_t$ does not depend on t.*

3. *Stochastic continuity: $\forall t \forall \epsilon > 0, \lim_{h \to 0} \mathbb{P}(|X_{t+h} - X_t| \geq \epsilon) = 0$.*

This last property does not mean that the sample paths are continuous, it merely excludes jumps at fixed times.

One example of such a model is the Black-Scholes model [2].

$$\frac{dS}{S} = mdt + \theta dW, \tag{32}$$

where S is the asset price, m is the average rate of growth and θ is a measure for the standard deviation of the returns called the volatility. $W(t)$ is a Wiener process, a continuous time stochastic process with independent increments $W(t) - W(s) \sim N(0, t - s)$.

4.1.2   Risk-neutral Valuation Theorem

One of the important assumptions in pricing options is the concept of an arbitrage free market. Basically this means there are no opportunities to make a risk-free profit better than the interest rate r of a bank account. If such an opportunity would exist, the market is assumed to immediately adjust prices to eliminate it before it can be used.

Mathematically this translates to the existence of a so-called risk-neutral measure $\mathbb{Q}$ which is equivalent to the real-world measure $\mathbb{P}$, under which the asset price divided by the bank account is a martingale. (A martingale is a stochastic process in which the conditional expectation of its value at time $T > t$, given the history of the process up to time t is the value of the process at time t.)

If we transform (32) to this measure, it would change to

$$\frac{dS_{\mathrm{RN}}}{S_{\mathrm{RN}}} = rdt + \theta dW_{\mathrm{RN}} \tag{33}$$

The change of measure changes the expected return to the return of the bank account and the random variable dW to a new random variable dW_{RN}.

This gives us a model for the risk-neutral asset price.

Under the risk-neutral measure we can write for the price $V(S_t, t)$ of a European option with expiration time T:

$$V(S_t, t) = e^{-r(T-t)} \mathbb{E}_t^{\mathbb{Q}}[V(S_T, T)]. \tag{34}$$

Using the definition of the expectation we can write this as

$$V(S_t, t) = e^{-r(T-t)} \int_{-\infty}^{\infty} V(S_T, T) f(S_T | S_t) dS_T, \tag{35}$$

23

where $f(S_T|S_t)$ is the transition density (under $\mathbb{Q}$) for the asset price.

For example, if we would consider a European option to buy an asset at time T for strike price E, the value of the contact at time T would be

$$V(S_T, T) = \max(S_T - E, 0) \tag{36}$$

If the asset price is modeled by (32) the transition density under $\mathbb{Q}$ would be

$$f(S_T|S_t) = \frac{1}{S_T\sqrt{2\pi\sigma^2(T-t)}} e^{-\frac{[\log\frac{S_T}{S_t} - (r - \frac{1}{2}\sigma^2)(T-t)]^2}{2\sigma^2(T-t)}} \tag{37}$$

Using this method we have to calculate the transition density and to solve the integral (35) to calculate the option prices. This is the basis for Transform methods in general and the CONV method in particular.

4.1.3 The CONV method

Transform methods aim at solving the risk-neutral valuation formula by means of Fast Fourier Transform (FFT) techniques. We will use the following notation for the continuous Fourier transform:

$$\mathcal{F}\{F(x)\} = \int_{-\infty}^{\infty} e^{-iux} F(x)dx \triangleq \hat{F}(u) \tag{38}$$

and its inverse:

$$\mathcal{F}^{-1}\{\hat{F}(u)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ixu} \hat{F}(u)du \tag{39}$$

The transform methods use log-transformed asset prices:

$$s = \log(S_T), \qquad k = \log(E) \tag{40}$$

and the risk-neutral density function $f(s)$ of the transformed stock price s.

The CONV method derives a pricing formula from the risk-neutral valuation formula by assuming

$$f(y|x) = f(y - x) \tag{41}$$

This assumption is full-filled for all processes with independent increments, such as Lévy processes.

For simplicity of notation we will use $x = S_t$ and $y = S_T$ throughout the derivation. This way (34) is written as

$$V(t, x) = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}_{t,S_t}[V(T, y)] \tag{42}$$

Using the definition of expectation we arrive at

$$V(t, x) = e^{-r(T-t)} \int_{-\infty}^{\infty} V(T, y) f(y|x)dy \tag{43}$$

24

At this point we use the assumption (41) and introduce a change of variables $z = y - x$:

$$e^{r(T-t)}V(t, x) = \int_{-\infty}^{\infty} V(T, x + z)f(z)dz \qquad (44)$$

Equation (44) is represented as a convolution integral here, a form suitable for a fast Fourier transform. The problem is that the Fourier transform may not exist. We dampen $V(t, x)$ by $e^{\beta x}$ to have $\int_{-\infty}^{\infty} |e^{\beta x}V(t, x)|dx < \infty$. This way the existence of the transform is ensured.

The CONV method is now defined by

$$
\begin{aligned}
& e^{r(T-t)}\mathcal{F}\{e^{\beta x}V(t, x)\} \\
=& e^{r(T-t)}\int_{-\infty}^{\infty} e^{iux}e^{\beta x}V(t, x)dx \\
=& \int_{-\infty}^{\infty} e^{iux}\left[\int_{-\infty}^{\infty} e^{\beta x}V(T, x + z)f(z)dz\right]dx \\
=& \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iux+\beta x}V(T, x + z)f(z)dzdx \\
=& \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} e^{i(u-i\beta)x}V(T, x + z)dx\right]f(z)dz \\
=& \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} e^{i(u-i\beta)(y-z)}V(T, y)dy\right]f(z)dz \\
=& \int_{-\infty}^{\infty} e^{i(u-i\beta)y}V(T, y)dy \int_{-\infty}^{\infty} e^{i(-u+i\beta)z}f(z)dz \\
=& \hat{V}_T(u - i\beta)\phi(-u + i\beta)
\end{aligned} \qquad (45)
$$

where $\phi(u)$ is the characteristic function of the density $f(x)$:

$$\phi(u) = \int_{-\infty}^{\infty} e^{iux}f(x)dx \qquad (46)$$

For Lévy processes this form can also be obtained via the Lévy-Khinchine formula.

To calculate the option price we apply the inverse Fourier transform:

$$e^{r(T-t)}V(t, x) = e^{-\beta x}\mathcal{F}^{-1}\{\hat{V}_T(u - i\beta)\phi(-u + i\beta)\} \qquad (47)$$

Again, this can be done efficiently by a fast Fourier transform due to the convolution in the expression.

This way we can efficiently calculate European option prices if we know the characteristic function of the risk-neutral density of the log-transformed stock price. For many stochastic processes the characteristic function is known in explicit form, while in some cases no explicit form for the density can be given.

4.1.4 Pricing American options

For the purpose of calibration, the only option types we consider are simple call and put options. The CONV method gives us a way to calculate European options. However, the bulk of the options traded in the market are American options. In this section we will show how the prices of Bermudan and American options can be derived using the algorithm for the European option price.

To price Bermudan options, the algorithm for European options is used repeatedly for each exercise point backwards in time. If there are M exercise times, we indicate them as $T_1, \ldots, T_M$.

Given the final pay-off function $V(t, x)$ and a given value of $S_{T_{M-1}}$, we can calculate $V_{\mathrm{HOLD}}(T_{M-1}, S_{T_{M-1}})$. This is the value of the contract if we choose not to exercise it at time T_{M-1} given the asset price $S_{T_{M-1}}$. However, as we have the option of exercising the contract at this time, the real price of the contract must be the maximum of holding the option and exercising it.

$$V_{\mathrm{REAL}}(T_i, S_{T_i}) = \max\{V_{\mathrm{HOLD}}(T_i, S_{T_i}), V(T_i, S_{T_i})\} \tag{48}$$

By calculating V_{REAL} for a range of values of S_{T_i}, we can approximate $V_{\mathrm{REAL}}(T_i, x)$ and use this approximation to calculate $V_{\mathrm{HOLD}}(T_{i-1}, S_{T_{i-1}})$ for a given value of $S_{T_{i-1}}$ by applying the CONV method.

From $V_{\mathrm{REAL}}(T_1, x)$ we can eventually calculate the option price $V(t, S_t)$, for a given $t < T_1$ and S_t using the CONV method.

We can approximate the American option price (which has unlimited exercise points) by a Bermudan option with many exercise points. As the number of exercise points increases, the Bermudan option price will approximate the American option price with increasing accuracy. However, this may become very costly as numerical integration should take place for each exercise point. To speed up this process, we use Richardson interpolation applied to the Bermudan approximation with 64, 32 and 16 exercise points.

4.2 Asset price models

As stated in section 4.1.3, we can evaluate option prices based on models of asset prices as long as we know the characteristic function of the risk-neutral density of the log-transformed stock price.

In this section we will review some asset price models and their characteristic functions. The parameters we want to estimate are written as $\theta_1, \ldots, \theta_k$ in each of these models.

4.2.1 Black-Scholes model

A well known model for asset prices is the Black-Scholes model [2]. The relative change in the asset price is modeled by a linear rate of return com-

bined with Brownian motion with a fixed standard deviation. In formula form this looks like

$$\frac{dS}{S} = mdt + \theta dW,$$ (49)

where S is the asset price, m is the average rate of growth and θ is a measure for the standard deviation of the returns called the volatility. $W(t)$ is a Wiener process.

When we transform this process to a risk-neutral measure the transformed price should be a martingale compared to the bank-account modeled by $\exp\{rt\}$, under the no-arbitrage assumption. This forces the drift to become equal to the instantaneous interest rate r. The risk-neutral model then reads

$$\frac{dS_{RN}}{S_{RN}} = rdt + \theta dW_{RN},$$ (50)

where S_{RN} is the risk-neutral asset price and dW_{RN} a Wiener process. If we denote $X = \ln S_{RN}$ we can apply Itô's formula to derive

$$dX = (r - \frac{1}{2}\theta^2)dt + \theta dW_{RN}$$ (51)

The characteristic function of X is now given by

$$\phi_{BS}(u,t) = \mathbb{E}[e^{iuX_t}] = \exp\{iu(r - \frac{1}{2}\theta^2)t - \theta^2u^2t\}$$ (52)

4.2.2 Variance Gamma model

The Variance Gamma (VG) [3] model adds more density to the tails of the normal distribution by replacing the deterministic time in the Black-Scholes model by a stochastic process based on a Gamma process. It causes the process to have infinitely many jumps in any time interval. The Variance Gamma process can be formulated as

$$X_{VG}(t;\theta_1,\theta_2,\theta_3) = \theta_1 G(t;\theta_3) + \theta_2 W[G(t;\theta_3)],$$ (53)

where θ_1 is the average rate of growth, θ_2 is the standard deviation, $G(t;\theta_3)$ is a Gamma process with mean rate unity and variance θ_3 and W is a Wiener process.

The characteristic function of the VG process is given by

$$\phi_{VG}(u,t,\theta_1,\theta_2,\theta_3) = \mathbb{E}\{\exp[iuX_{VG}(t)]\} = \left(\frac{1}{1 - i\theta_1\theta_3 u + \theta_2^2\theta_3 u^2/2}\right)^{t/\theta_3}$$ (54)

To model the asset price we use the process

$$S = S_0 \exp\{mt + P_{VG}\},$$ (55)

where m is a growth rate.

When we change the measure to a risk-neutral measure it changes to

$$S_{RN} = S_0 \exp\{(r + \omega)t + P_{VG,RN}\} \tag{56}$$

where ω is chosen so that $e^{-\omega t} = \phi_{VG}(-i, t)$, ensuring that the discounted risk-neutral process is a martingale.

To find the characteristic function we need, we denote $X = \ln S_{RN}$ and calculate its characteristic function as

$$\phi(u, t) = \exp\{iu(r + \omega)t\}\phi_{VG}(u, t) \tag{57}$$

4.2.3 CGMY model

The CGMY process [4] is a generalization of the VG process. It is derived from a different way of representing the VG model, i.e., as the difference of two Gamma processes:

$$X_{VG}(t; \theta_1, \theta_2, \theta_3) \stackrel{\text{law}}{=} G_p(t; \mu_p, \nu_p) - G_n(t; \mu_n, \nu_n) \tag{58}$$

Here we need the concept of the Lévy measure.

Definition 2 (Lévy measure). *Let $(P_t)_{t \geq 0}$ be a Lévy process on $\mathbb{R}^d$. The measure k on $\mathbb{R}^d$ defined by:*

$$k(A) = \mathbb{E}[\#\{t \in [0, 1] : \Delta P_t \neq 0, \Delta P_t \in A\}], \qquad A \in \mathcal{B}(\mathbb{R}^d) \tag{59}$$

is called the Lévy measure *of P: $k(A)$ is the expected number, per unit time, of jumps whose size belongs to A.*

From this representation the density of the Lévy measure of the VG process it can be shown to be

$$k_{VG}(x) = \begin{cases} \frac{\mu_n^2}{\nu_n} \frac{\exp(-\frac{\mu_n}{\nu_n}|x|)}{|x|} & \text{for } x < 0 \\ \frac{\mu_p^2}{\nu_p} \frac{\exp(-\frac{\mu_p}{\nu_p}|x|)}{|x|} & \text{for } x > 0 \end{cases} \tag{60}$$

The density of the Lévy measure of the CGMY process now is specified as

$$k_{CGMY}(x, \theta_1, \theta_2, \theta_3, \theta_4) = \begin{cases} \theta_1 \frac{\exp(-\theta_2|x|)}{|x|^{1+\theta_4}} & \text{for } x < 0 \\ \theta_1 \frac{\exp(-\theta_3|x|)}{|x|^{1+\theta_4}} & \text{fox } x > 0 \end{cases} \tag{61}$$

with $\theta_1 > 0$, $\theta_2 \geq 0$, $\theta_3 \geq 0$ and $\theta_4 < 2$.

The CGMY process $X_{CGMY}(t; \theta_1, \theta_2, \theta_3, \theta_4)$ is an infinitely divisible process of independent increments with a Lévy measure with density given by (61).

Its characteristic function is given by

$$\phi_{CGMY}(u, t, \theta_1, \theta_2, \theta_3, \theta_4) =$$
$$\exp\{t\theta_1 \Gamma(-\theta_4)[(\theta_3 - iu)^{\theta_4} - \theta_3^{\theta_4} + (\theta_2 + iu)^{\theta_4} - \theta_2^{\theta_4}]\} \qquad (62)$$

We model the asset price similar to the VG asset price:

$$S = S_0 \exp\{\mu t + X_{CGMY}\} \qquad (63)$$

with the risk-neutral process given by

$$S_{RN} = S_0 \exp\{(r + \omega)t + X_{CGMY,RN} \qquad (64)$$

where ω is chosen so that $e^{-\omega t} = \phi_{CGMY}(-i, t)$, again ensuring the discounted risk-neutral price process to be a martingale.

The characteristic function of the risk-neutral log-price $X = \ln S_{RN}$ is now given by

$$\phi(u, t) = \exp\{iu(r + \omega)t\}\phi_{CGMY}(u, t) \qquad (65)$$

4.2.4 Evaluation of the models

The main advantage of the Black-Scholes model is its simplicity. The entire option price is characterized basically by one parameter.

Under the model assumptions the market is complete: it is possible to replicate the option by constantly re-balancing a self-financing portfolio such that its value replicates the value of the option. This way there is no residual risk in writing a contract.

The main disadvantage of the model is its lack of realism. In the risk-neutral form, there is only one parameter that determines the whole behavior of the model. This would imply that if we would estimate the θ for various maturities and strike values we'd expect it to be constant. However, even when looking at the market prices for various strike prices with the same time to maturity, θ is not constant. It shows a curve that is known as the volatility smile or skew.

Additionally, the continuous sample paths may be satisfactory on a small scale, but if we look at real market data on a larger scale we see discontinuities that cannot be modeled.

Another weakness in models driven by Brownian motion concerns the tails of the distribution. Markets show jumps of multiple standard deviations too often for the distribution to be Gaussian and this excess kurtosis cannot be modeled in the Black-Scholes model.

Due to the jumps the VG and the CGMY models are better able to model the volatility smile for a given time to maturity. CGMY can capture it better since it has more parameters.

The main drawback of processes with discontinuities is the lack of the assumption of a complete market in these models. The change to the risk-neutral measure is not uniquely defined and corresponds to a different way of hedging. There is no way to construct a self-financing portfolio that will perfectly match a contingent claim at the time of maturity.

Though it is possible to "superhedge", that is, create a self-financing strategy that will almost always exceed the claim, even the cheapest possible superhedging strategy will be too expensive. To reduce the hedging cost to a realistic level, the option writer will have to take on some risk.

On the other hand, if such residual risk is the nature of the market, ignoring it by adapting a continuous model won't make it go away and it may be better to address the risk properly.

A second drawback is that the models don't perform as well for different maturities. The changes of time in the volatility smile are not reflected well in the model. This seems to originate from the assumption of independent increments in the stock price. Market data suggests a dependency in these increments.

4.3 Dividend payments

Twice a year, a company will pay out a certain amount of money to the shareholders as a share of the profits. As this will certainly affect the price of the asset, the option prices will be affected by this as well. The dividend payment will add a parameter δ to the model which is the percentage of the stock-price which will be payed out.

To model option prices with dividends, we use the Korn and Rogers model [5] without dividend announcements. While announcements are clearly more realistic, incorporating this approach in the CONV method is not feasible.

In this model, we do not model the stock price directly, but we model the stock price as the discounted expected values of all future dividend payments.

$$S_t = e^{rt}\mathbb{E}\left(\sum_{\{t_m > t\}}^{\infty} e^{-rt_m}D_m\right) \tag{66}$$

where D_m is the dividend paid at time t_m.

The dividend payments are modeled as the outcomes of an exponential Lévy process X multiplied by a constant λ:

$$D_j = \lambda X(t_j) \tag{67}$$

The Lévy process should satisfy

$$\mathbb{E}X_t/X_0 = e^{\gamma t} \tag{68}$$

for some $\gamma < r$. An additional assumption is that we have dividend payments at intervals $d_i > 0$.

$$t_m = md_i, \qquad m = 1, 2, \ldots \tag{69}$$

If we now write $T_d(t)$ as the first dividend payment from time t and $T_d^m(t)$ as the m-th payment after the first, we can write

$$
\begin{aligned}
S_t &= \sum_{m=0}^{\infty} e^{-r(T_d^m(t)-t)} \lambda \mathbb{E}_t X_{T_d^m(t)} \\
&= \sum_{m=0}^{\infty} e^{-r(T_d^m(t)-t)} \lambda X_t e^{\gamma(T_d^m(t)-t)} \\
&= \frac{\lambda X_t e^{-(r-\gamma)(T_d(t)-t)}}{1 - e^{-(r-\gamma)d_i}} \tag{70}
\end{aligned}
$$

This means that as a consequence of the model, the sizes of the dividend payments are random, but they are proportional to the price of the asset:

$$S(T_d) = \lambda X(T_d) \left(\frac{1}{1 - e^{-(r-\gamma)d_i}} - 1 \right) = S(T_d-) - \lambda X(T_d) = S(T_d-)e^{-(r-\gamma)d_i} \tag{71}$$

We will call this dividend percentage

$$\delta = 1 - e^{-(r-\gamma)d_i} \tag{72}$$

Here we see that λ does not influence the dividend payments, and we will choose its value to be $\lambda \equiv \delta$.

We can now compute the time-t price of a call option with strike K and expiry T by

$$e^{-r(T-t)} \mathbb{E} \left[(e^{-(r-\gamma)(T_d(T)-T)} X_T - K)^+ \right] \tag{73}$$

with $X_t = S_t e^{(r-\gamma)(T_d(t)-t)}$.

This means we can calculate the price by the CONV method if we apply it to

$$V(t,x) = (e^{-(r-\gamma)(T_d(t)-t)+x} - K)^+ \tag{74}$$

and characteristic function

$$\phi(u,t) = \phi_{\text{Lévy}}(u,t) \exp\{iu(\gamma + \omega)t\} \tag{75}$$

where ω satisfies $e^{-\omega t} = \phi_{\text{Lévy}}(-i, t)$ and $\phi_{\text{Lévy}}(u,t)$ is the characteristic function of a Lévy process, such as (54) or (62).

To estimate the first dividend that will be payed out we can write

$$
\begin{aligned}
\mathbb{E}_t D(T_d(t)) &= \mathbb{E}_t \lambda X_{T_d(t)} \\
&= \mathbb{E}_t \delta X_{T_d(t)} \\
&= \delta X_t e^{\gamma(T_d(t)-t)} \\
&= \delta S_t e^{(r-\gamma)(T_d(t)-t)} e^{\gamma(T_d(t)-t)} \\
&= \delta S_t e^{r(T_d(t)-t)}
\end{aligned}
\tag{76}
$$

5 Implementation

In this chapter we will first discuss the implementation of the Kriging algorithm. We will then examine the option pricing method in some detail. Finally we will look at the market data and propose an objective function.

5.1 Kriging

For the Kriging algorithm to work, we have to make a design decision for a stopping criterion in the genetic algorithm. A reasonable criterion seems to be to stop once the algorithm converges to the same optimum in 10 restarts in a row, with a maximum of 20000 function evaluations. This latter constraint prevents the algorithm to continue indefinitely in case there are several optimal points. In practice, the number of evaluations is usually smaller than 5000.

Using these values we will evaluate the effectiveness of adding the Maximum Variance Distance stopping criterion and the difference between using Latin Hypercube Sampling and Kriging Sampling.

We apply the algorithm to three different test problems: the 1-dimensional Rastrigin function, the 2-dimensional Schwefel function and the 2-dimensional Rosenbrock function. In all these problems we also examine the effect of having too few sample points to get a good initial impression of the function.

We will run the Kriging algorithm 100 times and check if it converged to the optimal point. We consider the algorithm to have converged if it is within 5% of the optimal point.

Based on these results we will then explain our design decisions.

5.1.1 1-dimensional Rastrigin function

The 1-dimensional Rastrigin function (Figure 5) is given by

$$f(x) = 10 + x^2 - 10\cos(2\pi x) \qquad x \in [-5.12, 5.12] \tag{77}$$
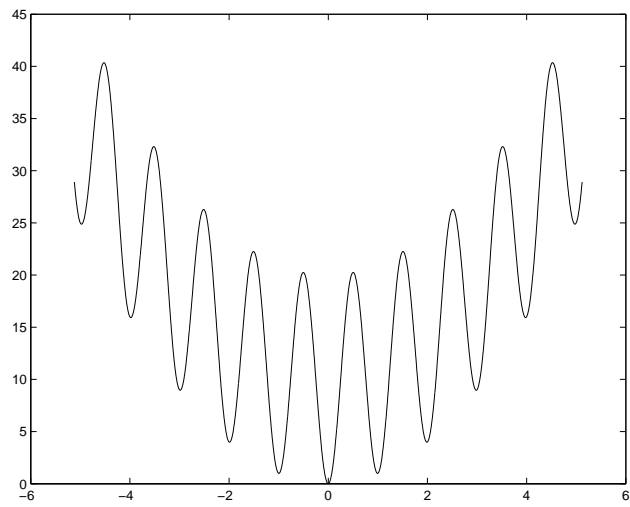
which has a minimum of 0 for $x = 0$.

Figure 5: 1D Rastrigin function

Num. start points	Sampler	MVD?	Num. converged	Av. num. steps
10	LHS	No	94	35.77
		Yes	98	35.02
	KS	No	82	27.00
		Yes	92	28.50
30	LHS	No	73	36.24
		Yes	57	40.61
	KS	No	100	39.37
		Yes	100	41.08

Table 1: 1D Rastrigin results

A surprising result in Table 1 is using LHS with 30 points which clearly leads to the worst results. The clustering of the LHS algorithm sometimes leads to a near singular R here and causes the resulting matrix multiplications with R^{-1} to produce some strange results for the predicted means and standard deviations, causing the algorithm to fail. Other than that, the algorithm seems to perform well, even with too few points to get a good starting approximation of the function. The differences between KS and LHS seem to arise partly from how many times a point was chosen close to the optimum. On the whole, adding MVD seems to be an improvement.

5.1.2 2-dimensional Schwefel function

The 2-dimensional Schwefel function (Figure 6) is given by

$$f(x,y) = -x\sin(\sqrt{|x|}) - y\sin(\sqrt{|y|}) \qquad x,y \in [-500, 500] \qquad (78)$$

which has a minimum of -837.9658 for $x = y = 420.9687$.
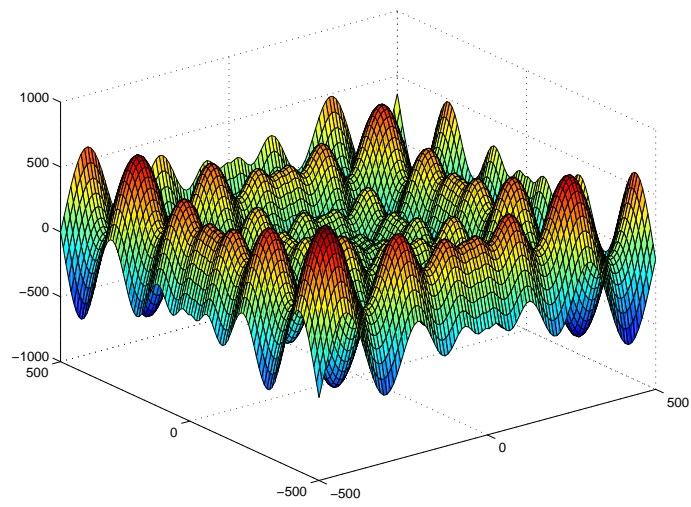
Figure 6: 2D Schwefel function

Num. start points	Sampler	MVD?	Num. converged	Av. num. steps
60	LHS	No	93	133.05
		Yes	95	129.21
	KS	No	82	133.10
		Yes	90	219.63
100	LHS	No	97	126.97
		Yes	100	128.88
	KS	No	82	201.16
		Yes	90	219.63

Table 2: 2D Schwefel results

In the runs in Table 2, LHS outperforms KS and MVD seems to add significantly to the number of times the optimum is found. However, it seems that in the LHS runs more often a point close to the optimum has been chosen as one of the start points which can be seen by the low average number of sample points before the algorithm considered itself converged.

5.1.3   2-dimensional Rosenbrock function

The 2-dimensional Rosenbrock function (Figure 7) is given by

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2 \qquad x, y \in [-2.048, 2.048] \qquad (79)$$

which has a minimum of 0 for $x = y = 1$.

Figure 7: 2D Rosenbrock function

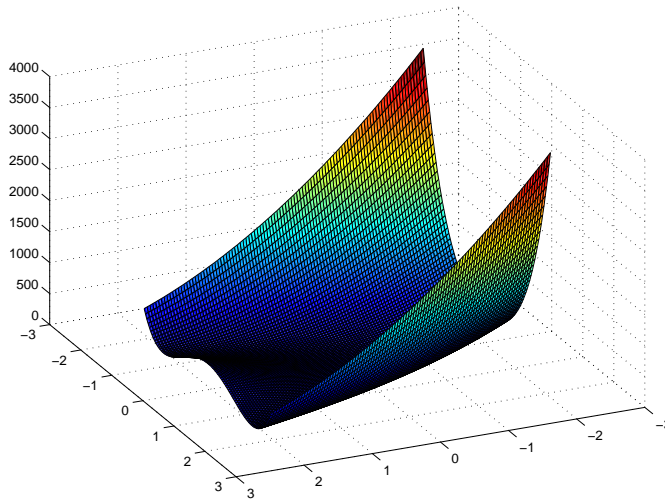| Num. start points	Sampler	MVD?	Num. converged	Av. num. steps
60	LHS	No	30	72.54
		Yes	77	80.10
	KS	No	62	96.06
		Yes	93	103.57
100	LHS	No	18	100.80
		Yes	21	101.69
	KS	No	60	126.76
		Yes	95	136.97

Table 3: 2D Rosenbrock results

For the Rosenbrock function the results in Table 3 are very clear: KS outperforms LHS and adding the MVD criterion adds a lot of accuracy for a small rise in the average number of steps needed. In the case of 100 start points we can see that how much the sampler affects the results. Sampling with LHS results in the algorithm considering itself converged almost immediately while the true optimum has not been reached yet. This is caused by the large valley where all the points have values close to the optimum. 100 start points are probably too many for this rather smooth function since only a small number of extra points are added in these tests.

5.1.4 Design decisions

Since we expect our market data objective to be a rather smooth function as well, we will use KS with the MVD criterion with 50 start points. To limit the duration of the algorithm we will stop the Kriging steps when we have 400 sample points and the MVD steps when we have 440 sample points.

5.2 Option pricing method

To implement the CONV method, we have to transform the problem to a discrete grid. This introduces the size of the grid 2^N on which to perform the Fourier transform. In the discrete model there is a trade off between the detail of the characteristic function (basically how many standard deviations to consider) and the step size in the payoff function both of which affect the accuracy of the algorithm. We choose a range of 8 standard deviations and estimate the maximal standard deviation to be 0.5. We use N as the sole parameter for the accuracy. Choosing specific values for each of the models will probably result in a lower value of N and thus faster calculations, but we will not go in to that in this thesis.

To get an indication which value of N to use for the various models we choose a parameter set for each model and calculate the call and put prices

of simple American options for 100 evenly spaced strikes between $K = 21$ and $K = 25$ for $N = 10, \ldots, 15$.

The parameter sets we use are shown in Table 4 are be reasonable approximations for the last data point from March 1th 2005 for ING options which expired in May 2005. We use a fixed interest rate of 2.3%.

Model	$(\theta_1, \ldots, \theta_n, \delta)$
BS	(0.1474, 0.0245)
VG	(-0.6374, 0.0584, 0.0468, 0.0268)
CGMY	(19.2618, 55.1580, 88.0093, 0.2328, 0.0244)

Table 4: Parameters used to assess the effect for N

To assess the error we compare the ℓ^∞-difference between the results for $N = 15$ with the others for each model.

Model	N	ℓ^∞ difference
	14	0.00001
	13	0.00002
BS	12	0.00009
	11	0.00031
	10	0.00099
	14	0.00370
	13	0.00110
VG	12	0.00680
	11	0.04770
	10	0.02600
	14	0.00001
	13	0.00004
CGMY	12	0.00040
	11	0.00087
	10	0.00730

Table 5: ℓ^∞ difference to $N = 15$

The results in Table 5 show a steady decline in ℓ^∞ error for Black-Scholes and CGMY. The results for Variance Gamma fluctuate but a downward trend is clearly visible.

Based on these results we choose $N = 13$.

5.3 Objective function

To compare different parameters in a model, we need an objective function which reflects how well a certain parameter set fits the market data.

S	K	$V_{\text{call}}^{\text{bid}}$	$V_{\text{call}}^{\text{ask}}$	$V_{\text{put}}^{\text{bid}}$	$V_{\text{put}}^{\text{ask}}$
	21.000	2.500	2.600	0.100	0.200
	22.000	1.600	1.700	0.300	0.350
23.460	23.000	0.850	0.900	0.600	0.700
	24.000	0.300	0.400	1.250	1.300
	25.000	0.100	0.150	2.050	2.150

Table 6: Example of market data

Market data is given by $(t_i, S_{t_i}, T_i, K_i, r_i, V_i^{\text{bid}}, V_i^{\text{ask}})_{i \in I}$ and whether it concerns a call or a put option. t_i is the current date, S_{t_i} is the current asset price, T_i is the expiry date, K_i the strike price, r_i is the instantaneous interest rate, V_i^{bid} and V_i^{ask} represent the bid-ask spread.

The data sets we have contain the value of the asset and the bid-ask spread for American call and put options. If more than one data set is available for one day, we will use the last. The data for one of these time points is presented in Table 6.

Given a model, we would like to find a parameter set θ such that for each $V_i^\theta = e^{-r_i(T_i - t_i)} \mathbb{E}^{\mathbb{Q}_\theta}[V(T_i, S_{T_i}, K_i)]$, we have $V_i^{\text{bid}} \leq V_i^\theta \leq V_i^{\text{ask}} \qquad \forall i \in I$ for both the call and the put options.

However, such an ideal parameter set may not exist in the model under consideration. In that case we would like to get the best fit in some sense.

We will use a weighted least square error with respect to the mid-market price $V_i^{\text{mm}} = \frac{V_i^{\text{bid}} + V_i^{\text{ask}}}{2}$.

$$g(\boldsymbol{\theta}) = \sqrt{\sum_{i \in I} c_i |V_i^\theta - V_i^{\text{mm}}|^2} \qquad (80)$$

where the c_is reflect our confidence in the individual data points $c_i = \frac{4}{|V_i^{\text{bid}} - V_i^{\text{ask}}|^2}$ as suggested in chapter 13.2 of [12].

The ideal parameter set θ^* is now defined as

$$\theta^* = \arg \min_{\mathbb{Q}_\theta \in \mathcal{Q}} g(\theta) \qquad (81)$$

The distance from the mid-market prices to the bid and ask prices is

now

$$
\begin{aligned}
\epsilon_0 &= \sqrt{\sum_{i \in I} c_i |V_i^{\text{bid}} - V_i^{\text{mm}}|^2} \\
&= \sqrt{\sum_{i \in I} c_i |V_i^{\text{ask}} - V_i^{\text{mm}}|^2} \\
&= \sqrt{\sum_{i \in I} \frac{4}{|V_i^{\text{bid}} - V_i^{\text{ask}}|^2} \left| \frac{V_i^{\text{bid}} - V_i^{\text{ask}}}{2} \right|^2} \\
&= \sqrt{\sum_{i \in I} 1}
\end{aligned}
\tag{82}
$$

where we use the data for both call and put prices.

Though this distance gives preference to points close to the mid-market prices, all solutions which fall between the bid and ask prices are equally valid. All solutions with an error greater than ϵ_0 will not have all their predicted values between the bid and ask prices, while solutions with a much smaller error will all be equivalent.

The main problem we now face is that the problem may be ill-defined. There may many solutions throughout the domain which all fulfill the requirement of lying between the bid and ask prices. It may be that there is no single stable solution.

This means that if we have found a θ^* and want to update it using new available data we might end up with a completely different θ^*!

6   Results

To examine the calibration algorithm, we use data sets for American options on ING around the dividend payment at April 28th 2005. During this period, the interest rate was approximately 2.3%. We look at those options which expired at the end of May 2005. The dividend was announced to be € 0.58 at February 17th 2005. Since the data sets we look at are all from after this date, we should be able to retrieve this value from our calibrations.

Instead of using the standard objective function we will use the logarithm of it's values as this "flattening" of the objective function helped the Kriging algorithm obtain better results.

We will look in detail at the calibration process from the data set of March 1st 2005. After this we will look at the stability of the optimal parameter sets we find.

6.1   Calibration

We will look at the calibration process for each asset price model in turn. To calibrate the models we will use the data set of March 1st 2005. This data set contains the asset price and the bid and ask spreads of 5 American call and put options as shown in Table 7.

S	K	$V_{\text{call}}^{\text{bid}}$	$V_{\text{call}}^{\text{ask}}$	$V_{\text{put}}^{\text{bid}}$	$V_{\text{put}}^{\text{ask}}$
	21.000	2.500	2.600	0.100	0.200
	22.000	1.600	1.700	0.300	0.350
23.460	23.000	0.850	0.900	0.600	0.700
	24.000	0.300	0.400	1.250	1.300
	25.000	0.100	0.150	2.050	2.150

Table 7: Data set of March 1st 2005

This number of options translates in a $\log(\epsilon_0) = 1.1512$.

Since we know the height of the dividend payment (€ 0.58) we would expect to find this value reflected in our calibrated model.

We will first examine the Black-Scholes driven model, then Variance Gamma and finally CGMY. Since each subsequent model has more parameters than the last, we would expect the errors to get smaller.

6.2   Black-Scholes

To examine our calibration method we will first use the Black-Scholes driven model. Since we have no extra information we will first use very broad ranges for the variables ($\theta \in [0.00, 0.50]$ and $\delta \in [0.00, 0.20]$ to see how the method responds.

Solution	log(error)	Num. points
(0.1470,0.0243)	1.1511	144
(0.1470,0.0203)	1.5733	55
(0.1470,0.0250)	1.1523	222
(0.1529,0.0203)	1.4369	146
(0.1509,0.0235)	1.1549	57
(0.1470,0.0250)	1.1523	164
(0.1490,0.0243)	1.1348	238
(0.1431,0.0235)	1.3442	70
(0.1490,0.0243)	1.1348	224
(0.1509,0.0235)	1.1549	157
(0.1431,0.0235)	1.3442	109
(0.1549,0.0211)	1.3489	71
(0.1509,0.0211)	1.3637	154
(0.1490,0.0258)	1.2248	195
(0.1490,0.0258)	1.2248	181
(0.1568,0.0219)	1.3396	171
(0.1450,0.0235)	1.2600	76
(0.1529,0.0219)	1.2632	65
(0.1529,0.0227)	1.2090	83
(0.1549,0.0219)	1.2877	184

Table 8: Calibration using BS with $\theta \in [0.00, 0.50]$ and $\delta \in [0.00, 0.20]$

As we can see in Table 8, the optima found lie in the same region, but the results still vary alot and there are only a couple of solutions with a log(error) smaller than $\log(\epsilon_0)$.

To gain more precise results we use the same method to investigate a smaller region, $\theta \in [0.10, 0.20]$ and $\delta \in [0.015, 0.030]$.

As we can see in Table 9 using this more reasonable search region immediately results in stable solutions. However, though the optimal solution has an error smaller than ϵ_0, it is still close to it, indicating that this model is not flexible enough to capture the curves of the call and the put prices with only these two parameters.

The optimal point for the Black-Scholes driven model is $(0.1490, 0.0242)$, corresponding to an estimated dividend payment of €0.572. In Figure 8 we show the prices for American options with various strike prices and add the bid and ask prices as dots to give an impression of the fit.

A common approach to get better results with Black-Scholes is to use different values of θ for different strikes. This would estimate the so-called volatility smile.

In an attempt to get a better fit with Black-Scholes, we tried to use

Solution	log(error)	Num. points
(0.1490,0.0242)	1.1348	54
(0.1486,0.0243)	1.1348	54
(0.1490,0.0242)	1.1347	53
(0.1490,0.0242)	1.1348	53
(0.1490,0.0242)	1.1348	53
(0.1486,0.0243)	1.1348	54
(0.1490,0.0243)	1.1348	54
(0.1486,0.0242)	1.1350	53
(0.1490,0.0243)	1.1348	53
(0.1490,0.0242)	1.1347	54
(0.1486,0.0242)	1.1350	53
(0.1486,0.0244)	1.1349	53
(0.1490,0.0242)	1.1347	53
(0.1490,0.0242)	1.1348	55
(0.1486,0.0242)	1.1355	53
(0.1490,0.0242)	1.1348	54
(0.1486,0.0244)	1.1349	52
(0.1486,0.0242)	1.1355	53
(0.1490,0.0244)	1.1352	53
(0.1490,0.0242)	1.1347	54

Table 9: Calibration using BS with $\theta \in [0.10, 0.20]$ and $\delta \in [0.015, 0.030]$

different values of θ for each of the strike prices for both call and put option and one value for δ. Unfortunately this did not improve the fit much as an improvement of the value of the call option made the fit for the put option worse and the other way around.
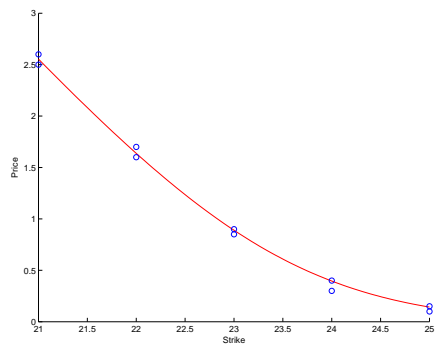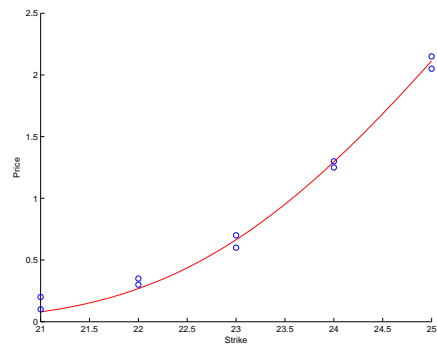
(a) Call option　　　　　　　　(b) Put option

Figure 8: Predicted option values for various strikes for Black-Scholes

6.3 Variance Gamma

As we have seen in the Black-Scholes driven model, it is important to have a good initial parameter range to get stable results from the calibration method. Since the Variance Gamma driven model has two more parameters (and therefore a much bigger search space) this is even more important here.

Since two of the variables in the Variance Gamma driven model have a link with the Black-Scholes driven model (θ_2 and δ), we will use those ranges from the Black-Scholes model.

Our initial search space will be $(\theta_1, \theta_2, \theta_3, \delta) \in [-0.50, 0.00] \times [0.10, 0.20] \times [0.01, 0.20] \times [0.015, 0.030]$.

Solution	log(error)	Num. points
(-0.1098,0.1521,0.1277,0.0249)	0.4346	86
(-0.1058,0.1521,0.1538,0.0252)	0.3587	441
(-0.1098,0.1501,0.1269,0.0257)	0.4309	441
(-0.1137,0.1509,0.1240,0.0253)	0.4270	441
(-0.1764,0.1380,0.1105,0.0269)	0.5134	73
(-0.3274,0.1380,0.0353,0.0255)	0.6396	96
(-0.1843,0.1384,0.1143,0.0271)	0.4550	441
(-0.1078,0.1509,0.1508,0.0254)	0.3531	441
(-0.1098,0.1517,0.1508,0.0255)	0.3601	441
(-0.2058,0.1447,0.0792,0.0244)	0.4901	86
(-0.1117,0.1505,0.1374,0.0257)	0.3982	441
(-0.1058,0.1529,0.1813,0.0252)	0.3257	441
(-0.1098,0.1509,0.1396,0.0256)	0.3828	441
(-0.1098,0.1525,0.1590,0.0253)	0.3549	441
(-0.1098,0.1509,0.1366,0.0254)	0.3850	419
(-0.1058,0.1521,0.1642,0.0253)	0.3325	441
(-0.1098,0.1509,0.1389,0.0257)	0.3881	441
(-0.1058,0.1533,0.1828,0.0248)	0.4020	114
(-0.1098,0.1509,0.1314,0.0254)	0.4013	441
(-0.1098,0.1509,0.1389,0.0255)	0.3805	441

Table 10: Calibration using VG with $(\theta_1, \theta_2, \theta_3, \delta) \in [-0.50, 0.00] \times [0.10, 0.20] \times [0.01, 0.20] \times [0.015, 0.030]$

As we can see in Table 10 the errors here are a lot smaller than in the Black-Scholes driven model. θ_2 and δ are reasonably stable. Most solutions have θ_1 around -0.11 and θ_3 around 0.15, though there are some solutions with a larger θ_1 and a smaller θ_3 but these solutions have a larger error.

To get more stable results we shrink our search space to $(\theta_1, \theta_2, \theta_3, \delta) \in [-0.15, -0.10] \times [0.12, 0.17] \times [0.12, 0.18] \times [0.024, 0.026]$

In Table 11 all the parameters are stable except θ_3 which varies between

Solution	log(error)	Num. points
(-0.1072,0.1519,0.1665,0.0254)	0.3227	441
(-0.1070,0.1521,0.1760,0.0253)	0.3147	441
(-0.1072,0.1521,0.1682,0.0253)	0.3197	441
(-0.1068,0.1521,0.1764,0.0254)	0.3142	441
(-0.1072,0.1521,0.1734,0.0254)	0.3155	441
(-0.1080,0.1517,0.1609,0.0254)	0.3287	441
(-0.1068,0.1523,0.1755,0.0253)	0.3151	441
(-0.1076,0.1517,0.1776,0.0257)	0.3292	441
(-0.1072,0.1519,0.1694,0.0254)	0.3185	441
(-0.1066,0.1521,0.1755,0.0254)	0.3148	441
(-0.1068,0.1521,0.1760,0.0254)	0.3147	441
(-0.1068,0.1521,0.1790,0.0254)	0.3142	441
(-0.1066,0.1521,0.1727,0.0254)	0.3166	441
(-0.1070,0.1521,0.1750,0.0254)	0.3146	441
(-0.1078,0.1519,0.1590,0.0254)	0.3299	441
(-0.1068,0.1523,0.1792,0.0253)	0.3143	441
(-0.1068,0.1523,0.1778,0.0254)	0.3141	441
(-0.1068,0.1521,0.1727,0.0254)	0.3160	441
(-0.1068,0.1521,0.1778,0.0254)	0.3141	441
(-0.1070,0.1523,0.1781,0.0254)	0.3142	441

Table 11: Calibration using VG with $(\theta_1, \theta_2, \theta_3, \delta) \in [-0.15, -0.10] \times [0.12, 0.17] \times [0.12, 0.18] \times [0.024, 0.026]$

0.1590 and 0.1790 but all those solutions give approximately the same error.

Unfortunately the method has broken off before convergence has been reached according to stopping criterion (23). To get some idea of how the method converges to the values in Table 11 we plot the running minimum and the value of the evaluated points in Figure 9 for all 20 runs.

We can see clearly the effect of adding some points by Maximum Variance Distance (which occurs after 400 points). There is a clear drop in the minimum errors of the evaluated points.

The optimal point found for the Variance Gamma driven model is $(-0.1068, 0.1521, 0.1778, 0.0254)$. This corresponds to an estimated dividend payment of € 0.598. Using these parameters we can calculate the values of American options for a range of strikes. This is shown in Figure 10 where the bid and ask prices are added as dots to give an impression of the fit.
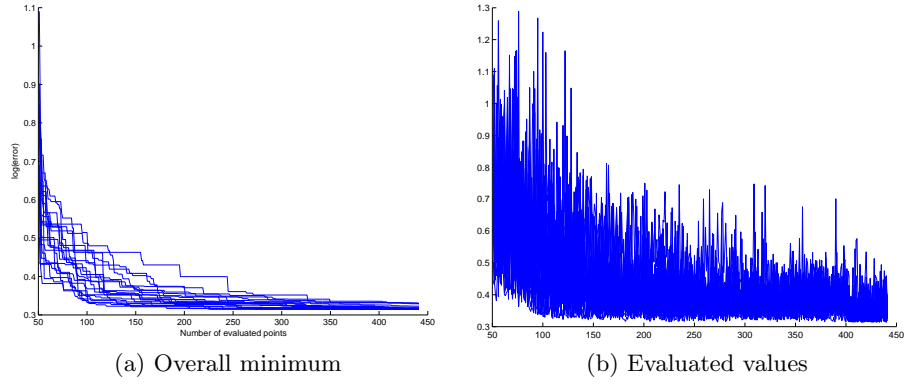
(a) Overall minimum   (b) Evaluated values

Figure 9: Analysis of the sample paths for VG
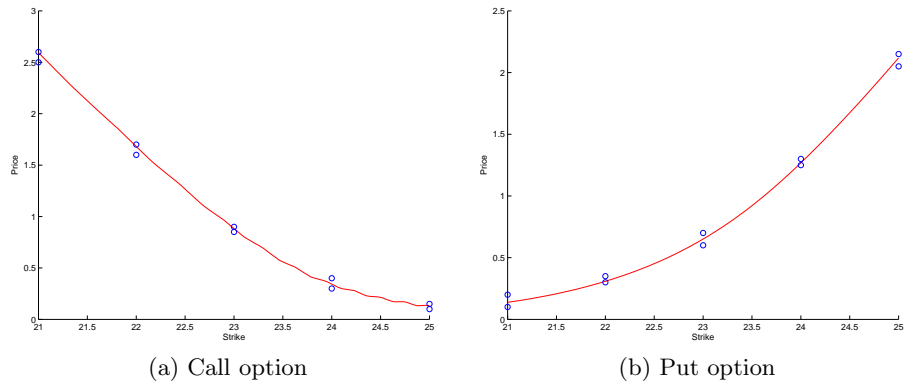

(a) Call option   (b) Put option

Figure 10: Predicted option values for various strikes for Variance Gamma

6.4 CGMY

To choose a search region for the CGMY driven model we use the solution from the Variance Gamma driven model. The CGMY model is the same as the Variance Gamma model for $\theta_1 = \frac{1}{\tilde{\theta}_3}$, $\theta_2 = \frac{1}{\sqrt{\frac{\tilde{\theta}_1^2 \tilde{\theta}_3^2}{4} + \frac{\tilde{\theta}_2^2 \tilde{\theta}_3}{2} - \frac{\tilde{\theta}_1 \tilde{\theta}_3}{2}}}$, $\theta_3 = \frac{1}{\sqrt{\frac{\tilde{\theta}_1^2 \tilde{\theta}_3^2}{4} + \frac{\tilde{\theta}_2^2 \tilde{\theta}_3}{2} + \frac{\tilde{\theta}_1 \tilde{\theta}_3}{2}}}$ and $\theta_4 = 0$, where the $\tilde{\theta}$s are the parameters from the Variance Gamma model (for more details see [4]).

Using this method we choose are initial search region as $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [5, 10] \times [10, 30] \times [20, 40] \times [-0.50, 0.50], [0.015, 0.030]$.

Solution	log(error)	Num. points
(8.0196,14.8627,24.1568,-0.3039,0.0287)	0.6380	441
(6.5294,24.9803,39.4509, 0.1745,0.0247)	0.4769	441
(7.9803,18.3921,27.7647,-0.1235,0.0255)	0.3182	441
(5.4117,21.5294,39.4509, 0.1588,0.0267)	0.4836	441
(6.9411,21.4509,38.4313, 0.0764,0.0271)	0.4681	441
(6.8823,22.5490,33.0980, 0.0803,0.0262)	0.5403	441
(5.8627,14.7058,23.2941,-0.1509,0.0251)	0.5497	441
(5.5686,21.8431,36.3137, 0.1588,0.0253)	0.4572	441
(9.1764,16.2745,25.8039,-0.2921,0.0285)	0.5779	441
(7.3137,18.5490,27.9215,-0.0882,0.0254)	0.3222	441
(6.5294,16.9803,32.1568,-0.0882,0.0271)	0.7339	441
(7.6470,17.5294,26.7450,-0.1352,0.0252)	0.3491	332
(5.3725,19.3333,29.3333, 0.0843,0.0251)	0.3992	295
(8.3921,23.6470,35.2156,-0.0254,0.0252)	0.6032	441
(9.7058,23.2549,34.3529,-0.0529,0.0248)	0.6134	441
(9.0980,21.3725,30.5098,-0.0843,0.0253)	0.3882	441
(6.0196,21.1372,31.1372, 0.0921,0.0252)	0.4073	441
(8.5294,15.5686,24.3137,-0.2568,0.0256)	0.5969	441
(6.3137,20.9019,36.7058, 0.0882,0.0270)	0.4684	441
(5.7450,24.8235,35.8431, 0.1980,0.0239)	0.6397	441

Table 12: Calibration using CGMY with $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [5, 10] \times [10, 30] \times [20, 40] \times [-0.50, 0.50] \times [0.015, 0.030]$

As we can see in Table 12, the solutions are very unstable. All solutions are much smaller than ϵ_0, but they are all larger than the minimal error we got for Variance Gamma.

Using the solutions with the smallest errors (log(error) < 0.41), we shrink our search region to $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [5, 10] \times [19, 22] \times [26, 32] \times [-0.15, 0.10] \times [0.025, 0.026]$.

The results in Table 13 show a big improvement in the errors. They are

Solution	log(error)	Num. points
(9.4705,20.6000,29.8117,-0.1156,0.0252)	0.3301	441
(8.3137,19.3176,28.7058,-0.1107,0.0256)	0.3200	441
(8.8823,19.5411,28.8470,-0.1254,0.0257)	0.3177	441
(8.8235,19.4823,28.5411,-0.1254,0.0251)	0.3079	441
(5.9803,19.1764,28.5176,-0.0294,0.0254)	0.3257	441
(9.1568,19.1647,28.4705,-0.1490,0.0258)	0.3124	441
(8.4705,19.0823,28.2117,-0.1235,0.0252)	0.2910	441
(8.9607,19.1058,28.3764,-0.1431,0.0255)	0.3062	441
(9.5686,20.0470,29.3176,-0.1352,0.0254)	0.3012	441
(9.2549,19.2941,28.5411,-0.1470,0.0255)	0.2985	441
(8.9019,19.5294,28.7294,-0.1274,0.0254)	0.2941	441
(8.6862,19.2117,28.4941,-0.1284,0.0254)	0.2956	441
(8.6862,19.3411,28.6588,-0.1254,0.0257)	0.3155	441
(8.2352,19.2823,28.3058,-0.1078,0.0251)	0.3174	441
(9.5294,20.8235,30.3529,-0.1107,0.0255)	0.3499	441
(8.2549,20.0470,29.3647,-0.0862,0.0254)	0.3242	441
(8.8039,19.0705,28.2588,-0.1372,0.0254)	0.2856	441
(8.7254,19.1529,28.3764,-0.1303,0.0254)	0.2955	441
(8.2156,19.2588,28.4235,-0.1058,0.0251)	0.3263	441
(9.4901,19.2823,28.4235,-0.1500,0.0252)	0.3497	441

Table 13: Calibration using CGMY with $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [5, 10] \times [19, 22] \times [26, 32] \times [-0.15, 0.10] \times [0.025, 0.026]$

now almost all below the level of Variance Gamma. However, the method still produces a lot of different solutions with approximately the same error.

In an attempt to improve the results we will shrink the search space even more to $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [8, 10] \times [19, 21] \times [28, 30] \times [-0.14, -0.10] \times [0.025, 0.026]$.

This last reduction of the search space gives results which are on average a bit better than the last step. The method does not yield one stable minimum, there is still a lot of variance in the parameters.

To analyze why the method breaks off we show the running minimum and the errors of the evaluated points for these 20 runs in Figure 11.

The effect of adding the last 40 points by Maximum Variance Distance is less clear here than in the case of the Variance Gamma driven model. This is probably an indication the method is further from convergence here.

The optimal point found by the method is (8.7764, 19.0235, 28.2196, −0.1381, 0.0253). This corresponds to an estimated dividend of €0.597. By calculating the prices of American options for a range of strikes and showing these in a plot with the bid and ask prices from the market data we can give an impression of the fit in Figure 12.

Solution	log(error)	Num. points
(8.7921,19.3372,28.5647,-0.1280,0.0254)	0.2930	441
(8.9411,19.3921,28.6039,-0.1326,0.0253)	0.2873	441
(8.2901,19.3764,28.6666,-0.1065,0.0254)	0.3086	441
(8.5960,19.4000,28.5647,-0.1194,0.0253)	0.2957	441
(8.1882,19.0705,28.2901,-0.1123,0.0254)	0.2944	441
(8.8313,19.3294,28.5019,-0.1298,0.0254)	0.2872	441
(9.4039,20.1529,29.3411,-0.1261,0.0254)	0.3095	441
(8.9333,20.1294,29.3803,-0.1101,0.0254)	0.3151	441
(8.4000,19.1254,28.3372,-0.1191,0.0254)	0.3026	441
(8.3450,19.1411,28.3137,-0.1160,0.0254)	0.2919	441
(8.0862,19.1882,28.2196,-0.1037,0.0252)	0.3125	441
(8.5411,19.3294,28.6745,-0.1192,0.0254)	0.3044	441
(8.9019,19.1882,28.1647,-0.1362,0.0251)	0.3021	441
(8.7764,19.0235,28.2196,-0.1381,0.0253)	0.2823	441
(8.6823,19.2274,28.4000,-0.1277,0.0253)	0.2869	441
(8.1647,19.2509,28.4156,-0.1067,0.0252)	0.3020	441
(8.9254,19.3450,28.4313,-0.1340,0.0251)	0.3000	441
(9.4745,19.9254,29.2392,-0.1365,0.0254)	0.3009	441
(9.0588,19.4235,28.5803,-0.1371,0.0252)	0.2976	441
(9.2313,19.5333,28.6901,-0.1400,0.0253)	0.2869	441

Table 14: Calibration using CGMY with $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [8, 10] \times [19, 21] \times [28, 30] \times [-0.14, -0.10] \times [0.025, 0.026]$
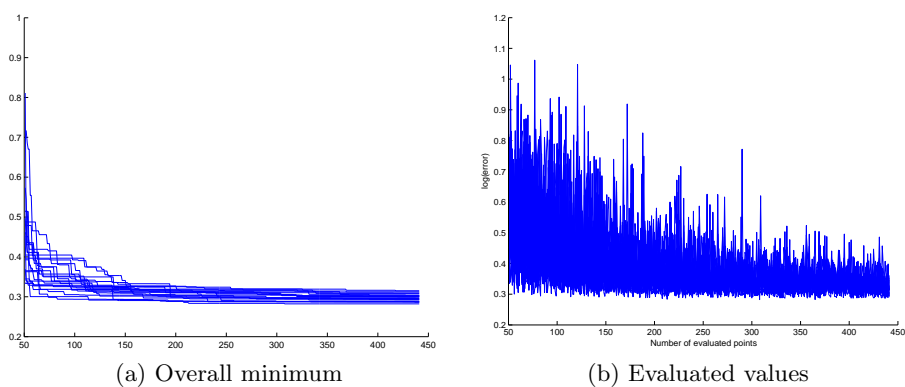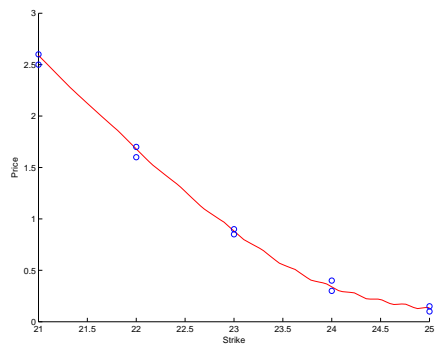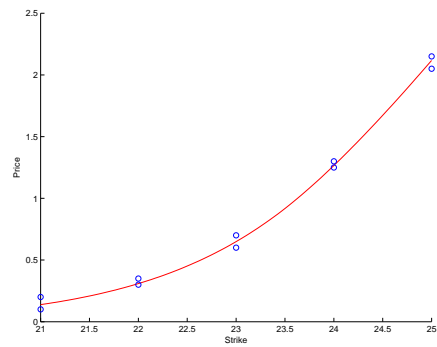

(a) Overall minimum   (b) Evaluated values

Figure 11: Analysis of the sample paths for CGMY

(a) Call option   (b) Put option

Figure 12: Predicted option values for various strikes for CGMY

6.5 Stability

In this section we will look at the stability of the solution we find. First we will look at how long the fit we found for the data of March 1st will remain a good fit for the data sets of the following days. After that we will calculate the optimal fit for the data set of March 8th and look at how much the parameters have changed.

6.5.1 Stability of the fit

In order to check how long the solutions we found stay valid, we check them against other data sets for put and call options for ING which expire at the end of May 2005.

Date	$\log(\epsilon_0)$	BS	VG	CGMY
01 03 05	1.1512	1.1347	0.3141	0.2823
02 03 05	1.1512	1.0554	0.3899	0.4841
03 03 05	1.1512	1.3206	0.2837	0.3630
04 03 05	1.1512	1.0605	0.6307	0.6641
07 03 05	1.1512	0.9544	0.8894	0.8996
08 03 05	1.2424	0.9008	0.5367	0.6612
09 03 05	1.2424	0.8750	1.2374	1.3195
10 03 05	1.2424	1.1517	1.2027	1.2629
11 03 05	1.2424	1.0664	0.9049	0.9445
16 03 05	1.2424	1.0834	0.8755	0.8853
17 03 05	1.2424	0.7546	1.5338	1.6165
18 03 05	1.2424	1.1106	1.0400	1.1938
21 03 05	1.2424	0.8915	1.3442	1.4808
22 03 05	1.2424	1.2547	1.2143	1.3129
23 03 05	1.2424	1.1876	1.1353	1.2452
24 03 05	1.2424	1.0228	1.4987	1.5686
25 03 05	1.2424	1.0387	1.6925	1.7806
28 03 05	1.2424	0.9370	1.5552	1.6963
29 03 05	1.2424	0.9555	1.3155	1.4233
30 03 05	1.2424	1.3745	1.6634	1.7599
31 03 05	1.2424	1.0706	1.5211	1.6456

Table 15: Stability over time

From Table 15 we can see that the fits we have found from the data of March 1st 2005 continues to give a good fit for one week and a reasonable fit for two weeks for Variance Gamma and CGMY. After that the fit becomes worse.

Surprisingly, while the fit of the Black-Scholes driven model is nowhere

really good, it declines less than the Variance Gamma and CGMY driven models.

6.5.2   Stability of the parameters

To check how much the parameters change over time, we calibrate Variance Gamma and the CGMY driven model to the data set of March 8th 2005 for American call and put options on the ING stock which expire at the end of May 2005. The value of the asset is € 23.730.

Using the same multi-level approach we start the search in the Variance Gamma driven model for $(\theta_1, \theta_2, \theta_3, \delta) \in [-0.50, 0.00] \times [0.10, 0.20] \times [0.01, 0.20] \times [0.015, 0.030]$. This results in a range of log(error) from -0.0904 to 0.3249. Shrinking the ranges to $[-0.20, -0.15] \times [0.14, 0.16] \times [0.08, 0.10] \times [0.024, 0.025]$ results in solutions which all lie closely around $(-0.1762, 0.1472, 0.0841, 0.0246)$ with a log(error) of -0.0939. This gives an estimated dividend payment of € 0.586.

Compared to the optimal solution for the data from March 1st $(-0.1068, 0.1521, 0.1778, 0.0254)$ we can see that only the second and the last parameter have stayed approximately the same.

For CGMY we start out with the search region $(\theta_1, \theta_2, \theta_3, \theta_4, \delta) \in [10, 15] \times [10, 30] \times [20, 40] \times [0.015, 0.030]$. This results in log(error) values ranging from -0.0860 to 0.4595. Zooming in on the best solution we examine the search space $[9, 11] \times [23.5, 25.5] \times [38, 40] \times [0.00, 0.05] \times [0.024, 0.026]$. This region yields a reasonably stable set of solutions around the optimal set of $(9.8078, 24.1231, 39.1607, 0.0476, 0.0246)$ with a log(error) of -0.0973. This also gives an estimated dividend payment of € 0.586.

Compared to the optimal set of March 1st $(8.7764, 19.0235, 28.2196, -0.1381, 0.0253)$ we can see that only the first and the last parameter stay relatively close to the former solution. Most notable is the change of sign of parameter θ_4 which changes the characterization of the CGMY process from finite activity for $-1 < \theta_4 < 0$ to infinite activity with finite variation for $0 < \theta_4 < 1$ (see [4]).

7 Conclusions

In this section we will discuss our experiences with the Kriging model in general, the different models and conclude with some thoughts on the estimated dividends.

7.1 Kriging

While Kriging seems to be a good method to find a minimum of an arbitrary function, it takes a large amount of time to determine which points to sample. This is mainly due to the many matrix multiplications (see (17 and 18)) needed in the genetic algorithm in the search for the next point to add. While there are few sample points this process is reasonably fast, but once there are a couple of hundred sample points the thousands of multiplications need a significant amount of time, and this increase is quadratic.

This increase in time forces us to limit the number of sample points after which we will break off the algorithm. In order to get good results, we have to make sure the initial Kriging approximation surface is reasonably close to the real surface. If the initial sample points do not give a good impression of the surface, the weights in the Kriging model will be off. This will in turn cause the model to have a bad guess at uncertainty between the sample points. If this uncertainty is too low, the model will stop the convergence too soon, maybe before the true optimum is found. If it is too high, the model will search too much in areas with few sample points instead of zooming in on the area where the points with low errors lie.

This first behavior can be seen in Table 8 where the Black-Scholes driven model is calibrated using a large range. The model considers itself converged on points near the optimum, but it does not converge to the optimum completely in most cases.

Finding a good initial guess for the parameter ranges is clearly very important. Using a multi-step approach to this problem works well enough, but is not time efficient.

7.2 Models

The Black-Scholes driven model is easy to calibrate since the space to be searched is small due to the small number of parameters. Unfortunately, the model does not fit the market data well enough.

The Variance Gamma driven model performs far better. If we use the multi-step approach we get good fits when we use one shrinking step to zoom in on the optimum of the first step.

The CGMY driven model also gives us good fits, even a little better than the Variance Gamma fits, but it is much harder to get consistent results. The initial sample points do not give a good enough fit when we use parameter

ranges that are too broad and the convergence is too slow to get good optima before the search is cut off. To get the results we did, we had to start out from a reasonable approximation using the Variance Gamma results. Though the fits we got were a bit better than the Variance Gamma fits, they showed comparable results over time.

The Variance Gamma driven model is the only model with good enough fits to be reliable and with few enough parameters to converge within the limits of our method if we use a multi-step approach.

7.3 Estimated dividend

If we look at the estimated dividends from the models, the predictions are close to, but not exactly on the mark. The thing to consider here is that the optimal solutions we found are not the best and only solutions to our problem. Every solution that produces option values between the bid and ask prices are equally valid, even though our error function gives preference to values closer to the mid-market price.

For instance, when we look at the values in the first step of the Variance Gamma driven model for the data set of March 1st (Table 10) we see that all these solutions will lie between the bid and ask prices, but have values of δ ranging from 0.0244 to 0.0271. This translates to range of estimated dividends from €0.574 to €0.638. This indicates that the bid-ask spread leaves too much room to estimate the dividend payment from just one data set.

References

[1] J. Sacks, W.J. Welch, T.J. Mitchell and H.P. Wynn (1989). Design and Analysis of Computer Experiments (with discussion). *Statistical Science* Vol 4, No. 4, 409-435

[2] F. Black and M. Scholes (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81 (3), 637-654

[3] D.B. Madan and E. Seneta (1990). The Variance Gamma (VG) Model for Share Market Returns. *Journal of Business* 63, 511-524

[4] P. Carr, H. Geman, D.B. Madan and M. Yor (2002). The Fine Structure of Asset Returns: An Empirical Investigation. *Journal of Business* 75, 305-332

[5] R. Korn and L.C.G. Rogers (2005). Stocks paying discrete dividends: modeling and option pricing. *Journal of Derivatives* 13 (2), 44-49

[6] R. Lord, F. Fang, F. Bervoets and C.W. Oosterlee (2007). A Fast and Accurate FFT-Based Method for Pricing Early-Exercise Options Under Lévy Processes. Working paper.

[7] S. Jeong, M. Murayama and K. Yamamoto (2005). Efficient Optimization Design Method Using Kriging Model. *Journal of Aircraft* 24 (2) 413-420

[8] K.V. Mardia and R.J. Marshall (1984). Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. *Biometrica* 71 (1) 135-146

[9] D.R. Jones, M. Schonlau and W.J. Welch (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* 13, 455-492

[10] M.D. McKay, W.J. Conover, R.J. Beckman (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, Vol. 21, No. 2, 239-245

[11] K. Krishnakumar (1989). Micro-Genetic Algorithms for Stationary and Non-Stationary Function Optimization. *SPIE: Intelligent Control and Adaptive Systems* Vol. 1196

[12] R. Cont and P. Tankov *Financial modelling with jump processes* Chapman & Hall/CRC, Boca Raton: 2004