J Glob Optim (2009) 43:219–230
DOI 10.1007/s10898-007-9224-3

Global optimization of higher order moments in portfolio selection

Dietmar Maringer · Panos Parpas

Received: 3 April 2007 / Accepted: 8 August 2007 / Published online: 22 September 2007

Abstract We discuss the global optimization of the higher order moments of a portfolio of financial assets. The proposed model is an extension of the celebrated mean variance model of Markowitz. Asset returns typically exhibit excess kurtosis and are often skewed. Moreover investors would prefer positive skewness and try to reduce kurtosis of their portfolio returns. Therefore the mean variance model (assuming either normally distributed returns or quadratic utility functions) might be too simplifying. The inclusion of higher order moments has therefore been proposed as a possible augmentation of the classical model in order to make it more widely applicable. The resulting problem is non-convex, large scale, and highly relevant in financial optimization. We discuss the solution of the model using two stochastic algorithms. The first algorithm is Differential Evolution (DE). DE is a population based metaheuristic originally designed for continuous optimization problems. New solutions are generated by combining up to four existing solutions plus noise, and acceptance is based on evolutionary principles. The second algorithm is based on the asymptotic behavior of a suitably defined Stochastic Differential Equation (SDE). The SDE consists of three terms. The first term tries to reduce the value of the objective function, the second enforces feasibility of the iterates, while the third adds noise in order to enable the trajectory to climb hills.

Keywords Portfolio selection · Heuristics · Global optimization · Markowitz model

Jel Classification C61 · G11

1 Introduction

Following the seminal work of Harry Markowitz [20], returns of financial assets are typically described by their mean, while risk is described by variance. Using the first two moments

D. Maringer
CCFEA, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK
e-mail: dmaring@essex.ac.uk

P. Parpas (✉)
Department of Computing, Imperial College, London SW7 2AZ, UK
e-mail: pp500@doc.ic.ac.uk

only is indeed sufficient if the investors have a quadratic utility function or the returns follow a Gaussian normal distribution. Unfortunately, neither condition holds in real life. Investors with non-increasing absolute risk aversion like positive skewness (see, e.g., [2]) as it indicates that extreme deviations from the mean tend to be on the plus side. Such investors dislike high kurtosis which indicates that extreme events have a high probability on either side. At the same time, stylized market facts indicate that higher order moments do matter as empirical data are skewed and, even more importantly, exhibit excess kurtosis and fat tails.

Against this backdrop, several approaches have been developed to incorporate higher order moments. In the model suggested by [4] with short-selling allowed and no risk-free asset, the efficient line in the mean-volatility space is a hyperbola. It is shown in [11] that the efficient set becomes a cone if skewness is considered in addition. Moreover an equilibrium asset pricing model was developed in [15] that includes mean, variance and skewness of returns. Their empirical evidence, however, is contradicted in [8]. In addition, there exist equilibrium models that include excess kurtosis (see, e.g., [26]).

In this paper we consider an extension of the celebrated Markowitz [20] model by incorporating the optimization of higher–order moments. The inclusion of higher order moments has been proposed as one possible augmentation to the model in order to make it more applicable (see e.g., [3,22], or [16]). The applicability of the model can be broadened by relaxing one of its major assumptions, i.e., that the rate of returns are normal. Given the numerical and theoretical challenges presented by these models we only consider a "vanilla" version of the problem where all other assumptions are kept.

Akin to the a mean variance framework, the selection problem can be split into two steps. From the universe of feasible portfolios, the majority can be classified as inefficient and should not be held by any investor for whom the usual assumptions of risk aversion apply. Which of the remaining efficient portfolios ought to be picked, however, depends on the investor's preferences; by means of a utility of preference function, the trade-off between the favorable expected return and positive skewness on the one side and the (unfavorable) variance and volatility on the other hand has to be found. This paper is concerned with the first of these two steps. In particular, we focus on a single period model and we propose two algorithms that can generate an efficient surface of portfolios. Every point on this surface will correspond to certain investor's preference selection.

For the problem described above we apply two promising approaches to stochastic global optimization. The first algorithm is Differential Evolution (DE). The algorithm was proposed in [32,33]. DE is a simple evolutionary method for continuous optimization problems which uses vectors to represent solutions. New solutions are generated in a two step procedure: First, three distinct individuals are selected, the first of these is used as the base vector to which the weighted difference of the remaining two is added; next, this combination is crossed-over with a fourth solution. Acceptance and replacement is based on evolutionary principles. Extended versions also include noise terms that generate diversity and hinder premature convergence. DE has recently been applied to constrained portfolio optimization in [19].

The second algorithm, which we call the diffusion algorithm, is based on a suitably defined Stochastic Differential Equation (SDE). The SDE consists of three terms. The first term indicates a descent direction for the objective function. The second term penalizes deviation from the feasible set (using a suitable definition of the Lagrange multipliers). If the trajectory passes through a local minimum then the first two terms will be zero. For this reason we consider a third term which adds some noise (using a Brownian motion term) to the trajectory and enables it to escape from local minima. The effect of the noise has to be large enough so that the the trajectory can escape all local minima, however its effects have to be gradually reduced in order to enable the algorithm to converge to the global minimum.

When the problem is unconstrained a similar algorithm was considered in [1,6,9,10]. When the problem has only linear constraints then an extension was proposed in [23]. The idea in [23] was to project the dynamics of the SDE onto the feasible set. The type of problems considered in this paper have non-linear constraints and so the method of [23] is not directly applicable. One of the aims of this paper is to discuss the use of an SDE approach to non-linearly constrained problems. Based partly on the numerical evidence given in this paper we believe that the diffusion method seems to be a promising approach for global optimization. We discuss such extensions through the lens of a highly interesting (and relevant) problem. The convergence of the algorithm is only discussed heuristically. A rigorous convergence proof is under development and will appear elsewhere [21].

The purpose of this paper is to show how stochastic optimization algorithms can be used to solve realistic financial planning problems. A review of applications of global optimization to portfolio selection problems appeared in [12]. A deterministic global optimization algorithm for a multi-period model appeared in [17]. This paper extends and complements the methods mentioned above in the sense that we incorporate the optimization of higher order moments into the model. The type of models we consider cannot, usually, be solved by deterministic algorithms. Consequently, practitioners are left with two options: solve a simpler, but less relevant model, or use a heuristic. In this paper we discuss two stochastic algorithms that represent very promising approaches to global optimization problems. Moreover, heuristic search methods have been found useful in different portfolio optimization problems including constraints on asset weights, distributions, or risk measures (see, e.g., [18]). We compare the performance of the two algorithms and discuss their merits in a real world application. Moreover, we discuss their practical implementation and tuning. Admittedly the model is rather simplistic. However, given the theoretical and computational difficulties involved with such models it is important to consider the simplified version of the problem in the hope that this approach will shed more light to the general case.

2 Higher moments in portfolio selection

To identify the portfolios amongst all feasible portfolios, the traditional optimization problem in a mean-variance framework can be stated as follows:

$$\min_x f(x) = \mathbb{V}(r_P) \tag{1a}$$

$$\text{s.t. } r_{P,t} = \sum_i x_i r_{i,t} \tag{1b}$$

$$\mathbb{E}(r_P) = r^* \tag{1c}$$

$$x_i \geq 0 \quad \forall i \tag{1d}$$

$$\sum_i x_i = 1 \tag{1e}$$

where $\mathbb{E}(\cdot)$ and $\mathbb{V}(\cdot)$ denote the mean and variance of portfolio returns, respectively. $r_{i,t}$ denotes the series of asset returns which, in this case, are input parameter.

Alternatively, the return constraint (1c) can be dropped and the trade-off between return and risk can be incorporated in a combined objective function, $f^{MV}(x) = \lambda\mathbb{V}(r_P) - (1-\lambda)\mathbb{E}(r_P)$. The scalar $\lambda \in [0, 1]$ represents the level of risk aversion: a risk neutral investor with $\lambda = 0$ will only maximize expected returns, while increasing values of λ put more weight on risk and therefore represent higher risk aversion. The efficient frontier can then be explored by

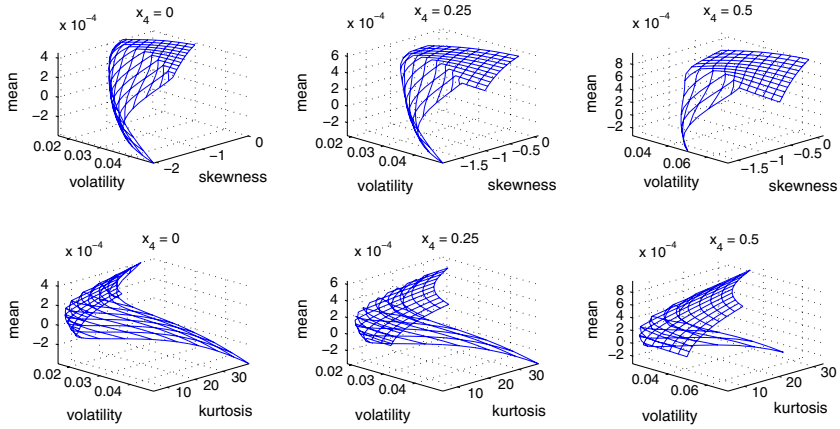

Fig. 1 Feasible sets for a four asset portfolio in the mean-volatility-skewness (top row) and mean-variance-kurtosis space (bottom row)

solving the problem for different values of λ. This remedies the problem that for a given value for the higher moment (skewness or kurtosis), the feasible range for the portfolio's expected return is not known in advance. The decision variables (x_i) represent the fractions of initial wealth invested in asset i. Note that this problem is a convex quadratic programming problem for which very efficient algorithms exist. The interested reader is referred to the reviews in [30] and [18] for more information regarding the Markowitz and related models.

If the returns exhibit higher order moments and if risk-averse investors do care about them, i.e., have a non-quadratic (and non-linear) utility function, then the definition of efficiency has to be extended. Akin to the original problem, an additional constraint can be introduced that ensures a certain level of skewness and kurtosis. In this paper, only one of the two higher moments is considered at a time. The new problem therefore reads as follows:

$$\min_x f(x) = \lambda \mathbb{V}(r_P) - (1-\lambda)\mathbb{E}(r_P) \tag{2a}$$

$$\text{s.t.} M(r_P) = M^* \tag{2b}$$

$$x_i \geq 0 \quad \forall i \tag{2c}$$

$$\sum_i x_i = 1 \tag{2d}$$

where $M(r_P)$ and M^* are the portfolio's actual and the target higher order moment (skewness or kurtosis), respectively.

The inclusion higher order moments makes the optimization problem non-convex and it can no longer be solved with the state of the art non-linear programming algorithms. The non-negativity constraint on the asset weights preclude analytic solutions. Figure 1 illustrates this problem for a 4 asset case. When one of the four weights is fixed (here: x_4) and the weights have to add to one (i.e., $x_3 = 1 - x_1 - x_2 - x_4$), the feasible set can be represented by a surface. A common approach is therefore to suggest modifications and simplifications to regain convexity. For the case of skewness, Konno et al. [14] consider the problem where the portfolio's skewness is maximized under constraints on the expected return and variance. After approximating the variance with the mean absolute deviation, the problem can be restated and solved by linear programming. Based on properties of the original problem, Boyle et al. [5] suggest local search method where a new portfolio is obtained that is close to base

portfolio which they find more efficient. Further related applications include [7,13,31] and [24].

3 Differential evolution

A recent addition to the class of population based heuristics is Differential Evolution (DE), suggested in [25,32,33]. In this method, each individual of the population represents one candidate which is encoded as a vector of all the decision variables (here: portfolio weights x). In each generation, for each individual p_0 a new solution p_n is generated in a two step procedure. First, a linear combination p_ℓ of three other, randomly picked individuals, p_1, p_2 and p_3 with $p_j \neq p_k \forall j \neq k$. For this new solution, p_1's solution, x^{p_1}, is chosen as the base vector to which the weighted difference of the other two solutions x^{p_2} and x^{p_3} is added; hence: $x^{p_\ell} = x^{p_1} + F \cdot (x^{p_2} - x^{p_3})$. Then, this solution x^{p_ℓ} is crossed over with x^{p_0} where with a probability ρ the i'th element of the vector comes from the p_0, otherwise from p_ℓ. When all new solutions for this generation have been produced, each of them is compared to their respective "parent" p_0 and replaces it if it outperforms it. In the course of generations, the individuals are expected to converge to the global optimum. Premature convergence can be avoided by adding noise terms z to the parameter F and/or to the difference vector, i.e., $x_i^{p_\ell} = x_i^{p_1} + (F + z_{1,i}) \cdot (x_i^{p_2} - x_i^{p_3} + z_{2,i})$. Further extensions suggest to include the elitist or several difference vectors in the creation of x^{p_ℓ} (see, e.g., [25]).

One of the advantages of DE is that it requires only a small number of technical parameters. In the basic version, the population size, the constant F and the cross over probability ρ have to be fixed; in the presented extended versions, the distributions for z_1 and z_2 have to be specified. Furthermore, these parameters require little tuning, and often standard values produce stable results. This was also found in preliminary experiments for this study, where the parameters were set to $F = 0.5$ and $\rho = 0.8$. However, it was also found for this problem that using both noise terms is beneficial; with a probability of 0.5, z_1 is a normally distributed variable with zero mean and 0.01 standard deviation, and zero otherwise. With a probability of 0.01, z_2 is a uniformly distributed variable in the range of $[-0.005; 0.005]$, and zero otherwise.

The non-negativity (2c) and budget constraints (2d) on the weights are satisfied with a repair mechanism where $x_i^{p_\ell} \leftarrow \max(0, x_i^{p_\ell})$ and $x_i^{p_n} \leftarrow x_i^{p_n} / \Sigma_j x_j^{p_n}$. For the constraint on the higher moments (2b), however, a punishment term is introduced that lowers the objective function by an amount reflecting the difference between targeted (M^*) and actual (M^a) value for the considered moment. More specifically, if the distance $\delta = |M^* - M^a|$ exceeds a certain small margin ϵ, a punishment term $\pi = (1 + \delta)\phi$ is introduced where ϕ is a scaling constant times the percentage of generations that has been passed already. Hence, the heuristic allows for exploration during the early generations, yet puts more pressure on constraint satisfaction during the latter. The population size is set to 50, and the number of function evaluations is limited to 100,000, resulting in 2,000 generations.

4 The diffusion algorithm

4.1 The algorithm

We describe the diffusion algorithm on the following general optimization problem:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g(x) = 0. \end{aligned} \tag{3}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, and $g : \mathbb{R}^n \to \mathbb{R}^m$, are assumed to be smooth. For the sake of argument, suppose that we did not have any constraints. A well known method for obtaining a solution to an unconstrained optimization problem is to consider the following Ordinary Differential Equation (ODE):

$$dX(t) = -\nabla f(X(t))dt. \tag{4}$$

By studying the behavior of $X(t)$ for large t, it can be shown that $X(t)$ will eventually converge to a stationary point of the unconstrained problem. A review of, so called, continuous-path methods can be found in [34]. In order to allow the trajectory to escape from local minima, it has been proposed by various authors (see [23] and references therein for a review of the literature) to add a stochastic term that would allow the trajectory to "climb" hills. One then considers the *diffusion process*:

$$dX(t) = -\nabla f(X(t))dt + \sqrt{2T(t)}dB(t) \tag{5}$$

where $B(t)$ is the standard Brownian motion in $\mathbb{R}^n$. It can be shown (e.g., [6,9,10]) that under appropriate conditions on f, and if the *annealing schedule* is chosen as follows:

$$T(t) \triangleq \frac{\Lambda}{\log(3+t)}, \quad \text{for some } \Lambda \geq c_0, \tag{6}$$

where c_0 is a constant positive scalar (the exact value of c_0 is problem dependent). Under these conditions, as $t \to \infty$, the transition probability of $X(t)$ converges (weakly) to a probability measure Π. The latter, has its support on the set of global minimizers. The problems considered in this paper are nonlinearly constrained. Consequently none of the above methods can be directly applied. In order to enforce the feasibility of the trajectory generated by an SDE, we propose to use the following:

$$dX(t) = -\nabla f(X(t)) - \nabla g(X(t))^T \lambda(X(t), t)dt + \sqrt{T(t)}dB(t). \tag{7}$$

where λ represent an estimate of the Lagrange multipliers of the problem and they are given by:

$$\lambda_i(x,t) \triangleq [\nabla g_i(x)\nabla g_i(x)^T]^{-1}[g_i(x) + T(t)\,\triangle\, g_i(x) - \nabla g_i(x)\nabla f(x)]. \tag{8}$$

$\nabla g_i(x)$ is used to denote the Jacobian of the i^{th} constraint. By $\nabla g(x) \in \mathbb{R}^{m\times n}$, we denote the $\mathbb{R}^{m\times n}$ Jacobian matrix associated with the constraints. $\triangle g_i(x) \in \mathbb{R}$ is used to denote the Laplacian of the constraints. Similar notation is used for the objective function. One way of studying the behavior of (7) is to first establish that that if t is large enough then $X(t)$ will eventually satisfy the constraints of the problem. Once this result is established then one needs to establish that the process will converge to the globally optimal solution of (3).The proofs are rather lengthy and will appear in a forthcoming paper [21].

4.2 Practical implementation

We note that the specification of the algorithm in the previous Section made use of exact gradients, Laplacians etc. Such data is not used in practice since noise is added to function evaluations, the benefits of exact gradient evaluations are wasted. For this reason all gradients were estimated using a gradient estimation approach proposed by Spall in [29]. Let Γ_t be a sequence of vectors in $\mathbb{R}^n$ and let each component of Γ_t be generated from a Bernoulli distribution with two equiprobable events $\{-1, +1\}$. For each function f whose gradient is required, we make two measurements:

$$f^+(X(t)) = f(X(t) + c_t\Gamma_t), \;\; f^-(X(t)) = f(X(t) - c_t\Gamma_t)$$

where c_t is a parameter to be described later. The gradient of f at $X(t)$ is then estimated by:

$$D(X(t)) = \left[\frac{f^+(X(t)) - f^-(X(t))}{2c_t\Gamma_{t1}}, \ldots, \frac{f^+(X(t)) - f^-(X(t))}{2c_t\Gamma_{tn}}\right]. \tag{9}$$

this gradient estimation is reminiscent of finite-differencing with the difference that only 2 measurements of the objective function are made in each iteration as opposed to $2n$. In our implementation we used: $c_t = \frac{c}{(t+1)^\gamma}$, where $\gamma = 0.14$.

From similar studies in the unconstrained case (e.g. [1,27]), we know that a deficiency of stochastic methods (of the type proposed in this paper) is that they require a large number of function evaluations. The reason for this shortcoming, is that the annealing schedule has to be sufficiently slow in order to allow the trajectory to escape from local minima. Therefore, whilst there are many sophisticated methods for the numerical solution of SDEs, we decided to use the cheaper stochastic Euler method. A further reason for not using higher order methods, such as Milstein's method, is that these methods require derivatives of the annealing schedule. While the functional form of this function has been identified, its direct use will slow down the algorithm too much. We will return to the topic of the annealing schedule later in this Section.

The stochastic Euler method is a generalization of the well known Euler method for ODEs to the stochastic case. Our implementation has two phases, in the first phase we generate a point based on the gradient estimate of the objective function, and we also add some noise. This step is given by:

$$X_r(t + \Delta t) = X(t) - Df(X(t))\Delta t + \sqrt{2T(t)\Delta t}u, \tag{10}$$

where Δt is the discretized step length parameter, and u is a standard Gaussian vector, i.e., $u \sim N(0, I)$. The second phase is a feasibility restoration phase, and it is an attempt to perform a numerical implementation of the Lagrange multiplier penalty term described in the previous section. This step is performed as in gradient restoration algorithms (see e.g., [28]). Given the X_r from (10) we attempt to compute a point $X(t + \Delta t)$ so that $g(X(t + \Delta t)) = 0$. Following [28] we compute a point $X(t + \Delta t)$ as follows:

$$X(t + \Delta t) = X_r(t + \Delta t) + \Delta X(t + \Delta t),$$

where

$$\begin{aligned} &\Delta X(t + \Delta t) = -Dg(X_r(t + \Delta t))\sigma(t) \\ &Dg(X_r(t + \Delta t))Dg(X_r(t + \Delta t))^T\sigma(t) = \zeta g(X_r(t + \Delta t)). \end{aligned} \tag{11}$$

In our implementation we solve the system in (11) for $\zeta = 1$, and then compute $X(t + \Delta t)$ as specified above. If the point obtained is feasible then we continue with phase one in (10). If the point is not feasible then we set $\zeta := \zeta/2$, and perform another restoration phase from the point where the previous restoration phase failed. This process is guaranteed to eventually yield a feasible point [28].

The algorithm starts by dividing the discretized time into k periods. Starting from a single strictly feasible point the algorithm generates m different trajectories. After a single period elapses, we remove the worst performing trajectory. Since all trajectories generate feasible points, we can assess the quality of the trajectory by the best objective function value achieved on the trajectory. We then randomly select one of the remaining trajectories, and duplicate it. At this stage we reduce the noise coefficient of the duplicated trajectory. When all the periods have been completed, in the manner described above, we count this event as one iteration. If the current incumbent solution vector remained the same for more than l iterations ($l > 4$,

in our implementation) then we reset the noise to its initial value. The algorithm terminates when the noise term is smaller than a predefined value ($0.1e-4$) or when after five successive resets of the noise term, no improvement could be made. In our implementation we used four trajectories, four periods, each of length $10e3$. The annealing schedule was started with the value $T(0) = 30$ and decreased by 0.7 in each iteration, i.e., $T(t+1) = T(t) \cdot .3$.

5 Empirical study

In this Section we report on the numerical performance of the two algorithms. We only studied the ten asset case to understand the effects of higher order terms to the portfolio selection problem. A similar approach can be used for larger problems.

5.1 Data

The empirical study is based on ten stocks included in the Dow Jones Industrial Average (DJIA). Using the adjusted daily prices downloaded from finance.yahoo.com for 2 March 2000 to 17 November 2006, 1684 log returns where computed. The top row of Fig. 2 contains scatter plots of the assets' returns mean and the other moments. Note that points on the same vertical level refer to the same asset. The mean returns are in the range of -4 to $+9$ basispoints, while daily standard deviations are mostly around 2%. The considered time period includes several turbulent market situations with extreme events. This leads not only to negative skewness in most of the included assets, but also to excess kurtosis and fat tails in all of the included assets. Not surprisingly, the assumption of a Gaussian normal distribution can be rejected for all of the assets both with a Jarque–Bera and a Kolmogorov–Smirnov test.

In the presence of higher order moments, optimizing with respect to mean and variance only can lead to highly undesirable effects. As the mean-variance optimization problem is (by definition) oblivious to skewness and kurtosis, the higher order moments of the portfolios become to some extent random. This can be seen from the graphs in the bottom row of Fig. 2

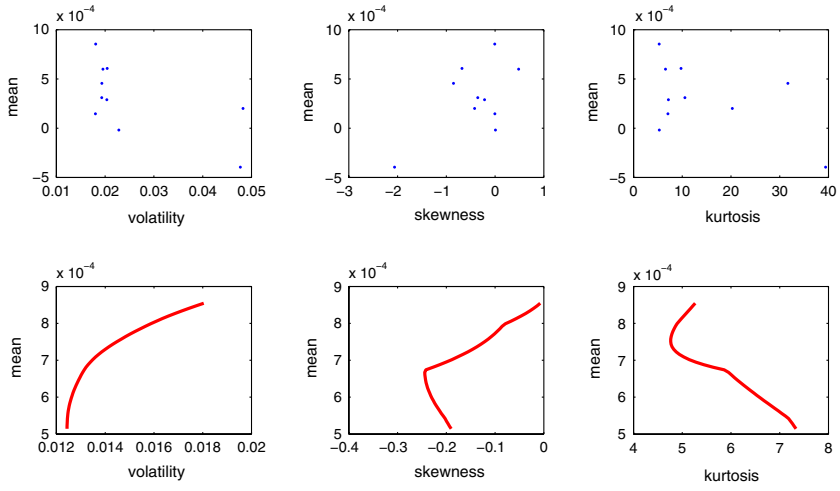

Fig. 2 Moments of stocks used for the computational study (top row) and mean-variance optimized portfolios (bottom row) for the 10 stock case

where points on the same vertical level represent the same portfolio. The portfolios are efficient in a mean-volatility space as they yield higher returns for any given level of volatility than individual stocks or other portfolios. However, low volatility portfolios come with higher kurtosis (lower half of the graphs), while higher volatility not only increase returns, but also skewness. In any case, the skewness is negative, and all the portfolios exhibit excess kurtosis; neither is desirable to an investor. Including these higher order moments is therefore a natural extension to the traditional optimization problem, and efficient sets are to be computed with respect to skewness and kurtosis.

5.2 Main findings

Other things equal, investors prefer higher return and skewness and lower volatility and kurtosis. Hence, when looking at three moments at a time and fixing one of them, efficient surfaces can be found. Figure 3 depicts the efficient lines with certain fixed skewness values. As can be seen, only portfolios with skewness below ca. -0.2 can be excluded as they are dominated. Above this value, increasing the skewness requires higher volatility; investors have therefore to find combinations that corresponds to their preferences the best. When projected into the mean-variance space (right), their different curvatures cause the lines to intersect. The hull of these projected lines constitute the Markowitz efficient line which therefore exhibits different skewness values for different mean-variance efficient portfolios (cf. Fig. 2). The same effect can be observed when the kurtosis; Fig. 4 depicts the efficient lines for different fixed values of kurtosis. Again, a projection onto the mean-variance space reveals that efficient lines for different values of kurtosis intersect, and the mean-variance efficient line is the hull. As in this case the kurtosis increases when the volatility is reduced, it becomes apparent that decision makers with a high level of risk aversion will not necessarily choose low volatility portfolios as they also exhibit a higher likelihood of extreme events. Which of the portfolios on the surface they should choose, however, depends again on their preferences and utility functions.

As was stated in the beginning of this Section, both methods can be applied to larger problems. We report some initial computational experience with such problems. Starting with the problem where only skewness is constrained initial results are reported in Table 1. The table shows the average number of function evaluations required to produce a point on the efficient frontier using the diffusion algorithm, and when skewness was constrained. The points were produced as described in Sect. 2. The statistics given in Tables 1 and 2 refer to the SDE algorithm. For the DE implementations, the number of function evaluations was

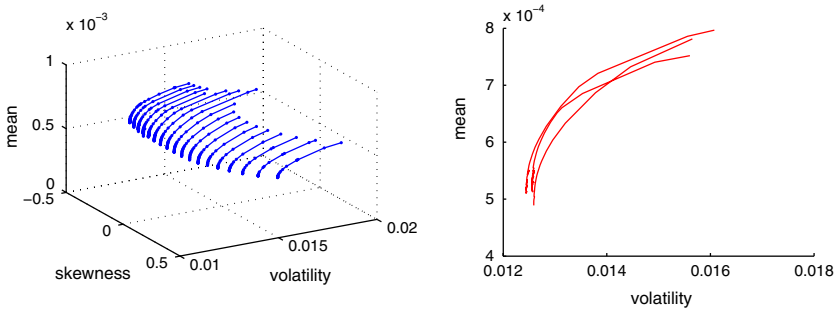

Fig. 3 Efficient lines for the 10 stock case with constraints on the skewness

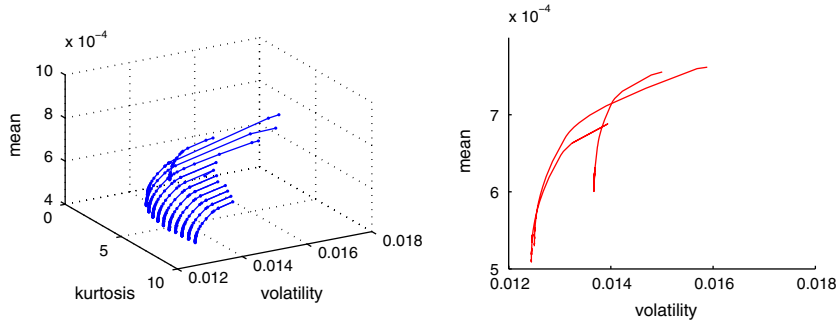

Fig. 4 Efficient lines for the 10 stock case with constraints on the kurtosis

Table 1 Diffusion method solution statistics—constraints on Skewness

Assets	Maximum Skewness	Average func. eval	Feasibility violation
10	0.498	5679	0
15	0.4985	24053	0
30	0.515	34655	0
45	0.561	56120	0
65	0.53	120332	0

Table 2 Diffusion method solution statistics—constraints on Kurtosis

Assets	Minimum Kurtosis	Average func. eval	Feasibility violation
10	4.30	6880	0
15	3.65	29960	0
30	3.64	31680	0
45	3.49	63136	0
65	3.42	159604	0

fixed to 100 000. The results are similar in terms of the optimal objective function values. Since the implementations were done on different platforms, however, using Linux and C++ for the SDE and Matlab on PCs with different processors, a comparison of CPU time is not informative.

For the model where kurtosis is constrained initial numerical results are reported in Table 2. An illustration of an efficient frontier of a problem with many assets and when the skewness is constrained is given in Fig. 5. While these results were obtained with the diffusion method, similar results can be obtained with the differential evolution algorithm.

Our numerical experience with the two algorithms has highlighted some advantages and some disadvantages for the two methods. A key advantage of DE is that it requires little tuning and can be used out of the box. Consequently a transparent implementation can be done requiring little input from the user other than the objective function, including a punishment term for constraint violations. On the downside it requires several restarts. For the problems considered in this paper ten restarts were performed, each restart requires more function evaluations than the SDE algorithm, typically 100,000 for the problems considered in this paper. The effectiveness of the method can be improved when DE is modified with

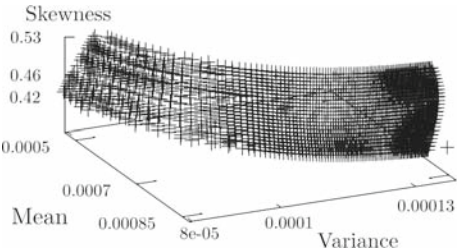

Fig. 5 An efficient frontier when skewness is constrained. This was generated with 65 assets and consists of about 2200 points

"noise" as described previously. With this modified version fewer restarts are required, but then some tuning is required. With respect to the diffusion method, a key advantage has been that the method always works with feasible solutions. This allows us to compare two different solutions directly. Moreover, the fact that the current solution is always feasible allows us to stop the algorithm prematurely but with an implementable strategy. A disadvantage of the method is that requires an annealing schedule for the reduction of the effects of the noise. This parameter is always difficult to fine tune. The scheme proposed here is simple, and works, but undoubtedly much more improvements can be made. If the problem has many local minima then the noise needs to be adopted accordingly, so it will probably be very difficult to obtain an annealing schedule that works in all conditions. Finally, the method borrows many ideas from nonlinear programming algorithms. This allows very efficient implementations of some of its aspects, e.g the gradient evaluation and projection steps.

6 Conclusions

We proposed an extension of the classical Markowitz model, to take into consideration higher order moments. We presented two stochastic algorithms for its solution and performed an empirical study. This work has highlighted many directions for future research. In terms of the modelling of the problem, we need to consider how to jointly enforce skewness and kurtosis constraints. In order to do this we need to identify how one choice for the level of one moment affects the other. In terms of the algorithms we need to make some improvements before we are able to apply these methods to larger problems. For future versions of the DE algorithm it will be beneficial to identify ways in which to better take advantage of "sophisticated guesses" for initial solutions. For the diffusion method more work on the annealing schedule might be needed, and ways to perform inexact projections will be developed.

Acknowledgment The second author gratefully acknowledges the support of EPSRC grant GR/T02560/01.

References

1. Aluffi-Pentini, F., Parisi, V., Zirilli, F.: Global optimization and stochastic differential equations. J. Optim. Theory Appl. **47**(1), 1–16 (1985)
2. Arditti, F.A.: Risk and required return on equity. J. Finance **22**(1), 19–36 (1967)
3. Athayde, G., Flores, R.: Incorporating skewness and Kurtosis in portfolio optimization: a multidimensional efficient set. In: Satchell, S., Scowcroft, A. (eds.) Advances in Portfolio Construction and Implementation, pp. 243–257. Butterworth-Heinemannpp, Oxford (2003)
4. Black, F.: Capital market equilibrium with restricted borrowing. J. Finance **45**(3), 444–455 (1972)
5. Boyle, P., Ding, B.: Portfolio selection with skewness. In Breton, M., Ben-Ameur, H., (eds.) Numerical Methods in Finance, GERAD Groupe d'études et de recherche en analyse des décisions. Springer (2005)

6. Chiang, T.-S., Hwang, C.-R., Sheu, S.J.: Diffusion for global optimization in $\mathbf{R}^n$. SIAM J. Control Optim. 25(3),737–753 (1987) ISSN 0363-0129.
7. Chunhachinda, P., Dandapani, K., Hamid, S., Prakash, A.: Portfolio selection and skewness: evidence from international stock markets. J. Bank. Finance **21**, 143–167 (1997)
8. Friend, I., Westerfield, R.: Co-skewness and capital asset pricing. J. Finance **35**(4), 897–913 (1980)
9. Geman, S., Hwang, C.-R.: Diffusions for global optimization. SIAM J. Control Optim. **24**(5),1031–1043 (1986) ISSN 0363-0129
10. Gidas, B.: The Langevin equation as a global minimization algorithm. In: Disordered systems and biological organization (Les Houches, 1985), volume 20 of NATO Adv. Sci. Inst. Ser. F Comput. Systems Sci. pp. 321–326 Springer, Berlin (1986)
11. Jean, W.H.: More on multidimensional portfolio analysis. J. Finance Quantit. Anal. **8**(3), 475–490 (1973)
12. Konno, H.: Applications of global optimization to portfolio analysis. In: Charles, A., Pierr H., and Gilles, J., (eds.) Essays and Surveys in Global Optimization, pp. 195–210. Springer (2005)
13. Konno, H., Suzuki, K.-I: A mean-variance-skewness portfolio optimization model. J. Operat. Res. Soc. Japan **38**(2), 173–187 (1995)
14. Konno, H., Shirakawa, H., Yamazaki, H.: A mean-absolute deviation-skewness portfolio optimization model. Ann. Operat. Res. **45**, 205–220 (1993)
15. Kraus, A., Litzenberger, R.H.: Skewness preference and the valuation of risk assets. J. Finance **31**(4), 1085–1100 (1976)
16. Lai, K.K., Yu, L., Wang, S.: Mean-variance-skewness-kurtosis-based portfolio optimization. First International Multi-Symposiums on Computer and Computational Sciences (IMSCCS'06) Vol. 2 , 292–297 (2006)
17. Maranas, C.D., Androulakis, I.P., Floudas, C.A., Berger, A.J., Mulvey, J.M.: Solving long-term financial planning problems via global optimization. J. Econom. Dynam. Control **21**(8-9), 1405–1425 (1997)
18. Maringer, D.: Portfolio management with heuristic optimization. Springer-Verlag (2005)
19. Maringer, D. Oyewumi, O.: Index tracking with constrained portfolios. To appear in Intell. Syst. Account. Finan. Manage
20. Markowitz, H.M.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
21. Parpas, P.: Constrained global optimization and stochastic differential equations. Technical Report, Department of Computing, Imperial College London (2007)
22. Parpas, P., Rustem, B.: Global optimization of the scenario generation and portfolio selection problems. In: Computational Science and Its Applications – ICCSA 2006, Proceedings Part III, vol. 3982 of Lecture Notes in Computer Science (2006)
23. Parpas, P., Rustem, B., Pistikopoulos, E.N.: Linearly constrained global optimization and stochastic differential equations. J. Global Optim. **36**(2), 191–217 (2006) ISSN 0925-5001
24. Prakash, A., Chang, C., Pactwa, T.: Selecting a portfolio with skewness: recent evidence from US, european, and latin american equity markets. J. Bank. Finance **27**, 1375–1390 (2003)
25. Price, K., Storn, R., Lampinen, J.: Differential Evolution: A Practical Approach to Global Optimization. Springer (2005)
26. Ranaldo, A., Favre, L.: How to price hedge funds: from two- to four-moment capm. Technical report, EDHEC Risk and Asset Management Research Centre (2003)
27. Recchioni M.C., Scoccia A. A stochastic algorithm for constrained global optimization. J. Global Optim **16**(3), 257–270 (2000) ISSN 0925-5001
28. Rom, M., Avriel, M.: Properties of the sequential gradient-restoration algorithm (SGRA). II. Convergence analysis. J. Optim. Theory Appl. **62**(1), 99–125 (1989) ISSN 0022-3239
29. Spall, J.C.: Introduction to Stochastic Search and Optimization. John Wiley & Sons, Hoboken, NJ (2003)
30. Steinbach, M.C.: Markowitz revisited: mean-variance models in financial portfolio analysis. SIAM Rev. **43**(1), 31–85 (electronic)(2001) ISSN 0036-1445
31. Stone, B.: A linear programming formulation of the general portfolio selection problem. J. Financial Quantit. Anal. **8**(4), 621–636 (1973)
32. Storn, R., Price, K.: Differential evolution — a simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report, International Computer Science Institute, Berkeley (1995)
33. Storn, R., Price, K.: Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. **11**(4), 341–359 (1997)
34. Zirilli, F.: The use of ordinary differential equations in the solution of nonlinear systems of equations. In: Nonlinear Optimization, 1981 (Cambridge, 1981), NATO Conf. Ser. II: Systems Sci., pp. 39–46. Academic Press, London (1982)