SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

On Portfolio Theory and Fractals

av

Ingvar Ziemann

2015 - No 20

On Portfolio Theory and Fractals

Ingvar Ziemann

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Yishao Zhou

2015

On Portfolio Theory and Fractals

Ingvar Ziemann

September 25, 2015

Abstract

We study optimal portfolio theory through a fractal framework in the presence of heavy tails and autocorrelated increments (Noah and Joseph effects). We show key results from the estimation of fractal dimensions and develop thereupon by proving the novel result that the Box-Counting dimension of a portfolio is concave. In order to illustrate the impact of the fractal dimension of return series a short exposition on fractional Brownian motion and Lévy stable processes is also rendered. We also introduce key concepts from optimization theory, portfolio theory and fractal geometry which are necessary to understand our approach, which to the best of our knowledge is new.

Acknowledgements

I wish to express my gratitude to Professor Y. Zhou at Stockholm University for supervising this thesis. Her feedback and ideas have been much appreciated while writing this thesis. Professor Zhou also sparked my interest in optimal portfolio theory and thus together with my readings of B. Mandelbrot allowed me to form the ideas for necessary this thesis. I would also like to thank Anu Kokkarinen for giving thorough feedback. Moreover, I would like to thank my father for introducing me to the writings of B. Mandelbrot, for the fruitful discussions we had and for helping me with MATLAB. Lastly, I wish to thank my mother for her eternal support.

Contents

1 Introduction

"Une même expression, dont les géomètres avaient considéré les propriétés abstraites, ... représente aussi le mouvement de la lumière dans l'atmosphère, quelle détermine les lois de la diffusion de la chaleur dans la matière solide, et quelle entre dans toutes les questions principales de la théorie des probabilités" [1].

- Joseph Fourier

This is not an essay on the theory of heat, as the quote by Fourier might indicate. Rather, it stands to elucidate the fact that subjects, at first glance seemingly different, may very well be united by their theoretical underpinnings. Its relevance comes from the fact that Mandelbrot's analysis of the length of the coast of Britain [2] is very much in the same spirit as the mathematics which may be applied to finance, or even probability. In common, they hold self-affine geometry; small parts resemble the whole. Our aim is to use traits of self-affinity found in finance to derive an alternative, or complementary, measure of the riskiness of a stock and apply the results to portfolio theory.

First, we need to understand what is meant by risk in the context of finance and economics. Economists have long equated, at least in their minds, the variance of some measurement quantity, most often price or returns, with risk. Perhaps the earliest account of such thinking is found in H. Markowitz's article *Portfolio Selection* of 1952 [3] even though economic awareness of the concept of risk in itself predates him. As early as in 1921 F. H. Knight published his seminal work titled *Risk, Uncertainty and Profit* [4] and its importance was realized even earlier. We will not go at great lengths to discuss what risk is, but informally note that any measure of risk should be increasing proportional to the probabilistic magnitude of our harm.

Bearing the concept of risk in mind, let us consider random processes. The tradition within financial literature is to consider normally and log-normally distributed prices or returns. This began as early as in 1900 when L. Bachelier published his doctoral dissertation [5]. The beauty of these processes lies in their predictability. In casual language, they exhibit what we call mild randomness.

Example 1.1 (Mild Randomness)**.** *Pick any one person at random (uniformly) from the earth's population and record their length. If we continue this process for a sufficient time we will expect the sample distribution to be approximately normal. Even though we expect some outliers their impact on the mean will eventually be negligible. In general, phenomena which are approximated by the various limit theorems of classical probability theory are considered mild and their distributions are characterized by flat tails.*

If we apply the same reasoning toward financial markets we would expect a rather calm set of price changes. Nevertheless, the data strongly refute any such claims.

Example 1.2 (Wild Randomness)**.** *Consider the stock market. If returns were normal we would expect outliers to be negligible to expectation. Between 1916 and 2003 theory would predict six days of index swings beyond* 4.5% *- there were 366 such days. Similarly, changes of* 7% *or more would occur once every 300,000 years, but we have seen 48 such days* [6].

Not only does variance as a model of risk underestimate the tails of the distribution of returns, as illustrated in example 1.2, but returns also demonstrate trend-like behavior known as momentum within the financial litterature [7]. Clearly, the implications for the individual investor are huge.

Since L. Bachelier, in the very beginning of the 20th century, normality assumptions have played a major part in modern finance. The perhaps two most important models in modern finance, the Markowitz portfolio model from [3] and the Black-Scholes-Merton stochastic differential equation from [8], rest heavily on the Gaussian foundation. Even though we by no means wish to refute their significance to the development of modern financial mathematics it is our belief that over-confidence in variance as a successful estimator of risk can, and has, had disastrous consequences. Especially since the second moment, variance, is infinite for a whole class of distributions known as Lévy stable distibutions (with the exception of the extreme case of the normal distribution). Hence, owing to B. B. Mandelbrot's ingenuity, we intend to expand on H. Markowitz's intellectual heritage with a fractal measure of risk. It is our ambition that such a measure better captures extreme variations and anti-persistent tendencies and thus better accounts for the risk of disaster, taken in the meaning of major crashes of any collection of assets.

2 Background material

In order to treat the ideas discussed in the introduction with logical clarity, a mathematically precise framework is required. Our intention is to render a self-contained account of the applications of fractals to portfolio theory and thus assume only the basics of real analysis (such as from [9]), set and measure theory (as presented in [10]), probability and topology - areas which, though the foundation of our study, are not of primary interest to us. We also require some knowledge of Fourier transforms such as from [11] or [12] to calculate dimensions of stochastic processes, especially due to its close relation with the characteristic function of a random variable. References for more difficult results used as background are given unless they are proven in the appendix. We now review concepts key to our purpose from optimization theory, portfolio theory, topology, the study of fractals, and their relation to probability.

2.1 Optimization

The aim of this section is to present optimization theory, specifically through the method of saddle point optimality. We have chosen this route as we believe it to be minimal in the sense that we do not have to delve particularly deep into convex analysis compared with the machinery needed for the alternate route of proving KKT necessary and sufficient conditions directly [13]. Nevertheless we are still able to provide a rigorous account of that which will be needed in treating the main problem. Further, the intrinsic geometric nature of a saddle point is appealing and easy to grasp without relying on a background in optimization theory. For a more general and extensive treatment of convex analysis and optimization we refer to [14] and [13].

2.1.1 Terminology

Before we commence we need the following intuitive definition of optimality.

Definition 2.1. *$\bar{x}$ is a local minimum of $f : X \to \mathbb{R}$ if $f(\bar{x}) \leq f(x)$ for all $x \in N_\varepsilon(\bar{x}) \cap X$ and for some $\varepsilon > 0$. Similarly, $\bar{x}$ is a global minimum of $f : X \to \mathbb{R}$ if $f(\bar{x}) \leq f(x)$ for all $x \in X$. We define maxima by calling $\bar{x}$ a local (global) maximum of $f : X \to \mathbb{R}$ if $\bar{x}$ is a local (global) minimum of $-f(x)$.*

In general, we consider a problem of the form below, where we look for the optimal solution. Moreover, it is clear from the above definition that, we can without loss of generality, solely study mimima.

Problem 1 (The General Primal Optimization Problem)**.**

$$\begin{aligned} &\min_{x \in X} f(x) \\ &s.t.\ \ g_i(x) \leq 0, h_j(x) = 0 \end{aligned} \tag{1}$$

where f, g_i, h_j for $i = 1, ..., m$ and $j = 1, ..., l$ are functions from $X \subset \mathbb{R}^n \to \mathbb{R}$. We can thus define the feasible set

$$S = \{x \in X : g_i(x) \leq 0, h_j(x) = 0\}. \tag{2}$$

Any member $x \in S$ is called a feasible solution.

Remark 2.1.1. *We will often write $g(x) = (g_1(x), ..., g_m(x))^t, h(x) = (h_1(x), ..., h_l(x))^t$.*

We now define convexity for functions and for sets.

Definition 2.2. *We say that a set X is convex whenever for all $x, y \in X$, $\lambda x + (1 - \lambda)y$ is also a member of X for all $\lambda \in [0, 1]$.*

A function $f : X \to \mathbb{R}$ for a convex set X is convex whenever

$$\lambda f(x) + (1 - \lambda)f(y) \geq f(\lambda x + (1 - \lambda)y) \tag{3}$$

for all $x, y \in X$ and $\lambda \in [0, 1]$. Similarly, f is concave whenever $-f$ is convex.

Remark 2.2.1. *If the inequality (3) is strict, we say that the function f is strictly convex (or concave when suitable).*

Example 2.1. *An important class of convex sets are the polyhedral sets, $S = \{x \in \mathbb{R}^n : Ax \leq b\}$. They are the finite intersections of closed half-spaces and are easily shown to be convex. Let $x, y \in S$ and $\lambda \in [0, 1]$. Then since $x, y \in S$, $Ax \leq b$ and $Ay \leq b$. Thus $A(\lambda x + (1 - \lambda))y = \lambda Ax + (1 - \lambda)Ay \leq b$, which precisely means that the convex combination also lies in S.*

We now present some propositions which make it much easier to verify that a function is convex. It is not just of theoretical interest to study which operations preserve convexity; it is in fact of great practical importance as it allows us to reduce checking convexity of more difficult functions into that of checking convexity for several more basic functions. Aside from the trivial operation, addition, convexity is also preserved under multiplication and composition of functions under suitable conditions. We will prove these statements in the following.

Proposition 2.1. *Let $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}$ be non-negative and convex on the convex set $X \subset \mathbb{R}^n$ and assume that $(f(x) - f(y))(g(x) - g(y)) \geq 0$ for all $x, y \in X$. Then fg is convex on X.*

Proof. Let $\alpha \in [0, 1]$ and $x, y \in X$. Let

$$F(x, y, \alpha) = \alpha(fg)(x) + (1 - \alpha)(fg)(y) - (fg)(\alpha x + (1 - \alpha)y). \tag{4}$$

Note that convexity of fg is equivalent to $F(x, y, \alpha) \geq 0, \forall \alpha, x, y$. Using convexity of f and g, one sees that

$$\begin{aligned} F(x, y, \alpha) &= \alpha(fg)(x) + (1 - \alpha)(fg)(y) - (fg)(\alpha x + (1 - \alpha)y) \\ &\geq \alpha(fg)(x) + (1 - \alpha)(fg)(y) \\ &\quad - (\alpha f(x) + (1 - \alpha)f(y)) (\alpha g(x) + (1 - \alpha)g(y)) \\ &= (\alpha - \alpha^2)(fg)(x) + (1 - \alpha - (1 - \alpha)^2)(fg)(y) \\ &\quad - \alpha(1 - \alpha)(f(x)f(y) + g(x)g(y)) \\ &= \alpha(1 - \alpha) (f(x)g(x) + f(y)g(y) - f(x)f(y) - g(x)g(y)) \\ &\geq \alpha(1 - \alpha)D(x, y) \end{aligned} \tag{5}$$

where $D(x, y) = (f(x) - f(y))(g(x) - g(y))$. This was assumed to be non-negative and thus the result follows. $\square$

Proposition 2.2. *Let $f : Y \to \mathbb{R}$, $g : X \to Y$ be convex functions, with f non-decreasing on the convex set $Y \subset \mathbb{R}$. Then $f \circ g$ is convex on X.*

Proof. Let $\alpha \in [0, 1]$. Since f is convex and non-decreasing and g is convex.

$$f(g(\alpha x + (1 - \alpha)y)) \leq f(\alpha g(x) + (1 - \alpha)g(y)) \leq \alpha f(g(x)) + (1 - \alpha)f(g(y)). \tag{6}$$

The result follows. $\square$

The algebraic definition of convexity is tedious to check even for simple sets and functions. We therefore present the following equivalent forms for suitably differentiable functions.

Proposition 2.3. *A differentiable function f on a non-empty open convex set X is convex if and only if it holds for all $x, y \in X$ that*

$$f(y) \geq f(x) + \nabla f(x)^t (y - x) \tag{7}$$

Proof. Let $f(x)$ be convex on X. Then for arbitrary $x, y \in X$ and $\alpha \in [0,1]$

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y) \leftrightarrow \frac{f((x + \alpha(y-x)) - f(x)}{\alpha} \leq f(y) - f(x). \quad (8)$$

Letting $\alpha \to 0$ this becomes $\nabla f(x)^t(y-x) \leq f(y) - f(x)$.

Let also $a, b \in X$, $\omega \in [0,1]$ and assume the inequality (7). Then

$$\begin{aligned}(1-\omega)f(a) + \omega f(b) &\geq f(x) + (1-\omega)[\nabla f(x)^t(a-x)] + \omega[\nabla f(x)^t(b-x)] \\ &= f(x) + \nabla f(x)^t[\omega(b-x) + (1-\omega)(a-x)] = f(x) + \nabla f(x)^t[(1-\omega)a + \omega b - x]\end{aligned} \quad (9)$$

and then letting $x = (1-\omega)a + \omega b$ gives

$$(1-\omega)f(a) + \omega f(b) \geq f(x) + \nabla f(x)^t(0) = f(x) = f((1-\omega)a + \omega b). \quad (10)$$

Thus we have proven both directions of the proposition.

□

We end this section with the following useful characterization of convexity. It possesses the added benefit of being easy to check for most functions of interest to us.

Proposition 2.4. *Let X be a non-empty open convex set and let $f : X \to \mathbb{R}$ be twice differentiable on X. Then f is convex on X if and only if the Hessian matrix of f is positive semidefinite (PSD) at each point of X.*

Proof. Let f be convex and let $y \in X$. Since X is open we find that for sufficiently small $|\lambda| \neq 0$ and any $x \in X$, $x + \lambda y$ is also in X. By proposition 2.3 and since f is twice differentiable

$$\begin{aligned}&f(y + \lambda x) \geq f(y) + \lambda \nabla f(y)^t x \text{ and} \\ &f(y + \lambda x) = f(y) + \lambda \nabla f(y)^t x + \frac{\lambda^2}{2} x^t H(y) x + \lambda^2 \|x\|^2 \mathcal{O}(y, \lambda x).\end{aligned} \quad (11)$$

Subtracting the latter from the former of (11), we get

$$\frac{\lambda^2}{2} x^t H(y) x + \lambda^2 \|x\|^2 \mathcal{O}(y, \lambda x) \geq 0. \quad (12)$$

Dividing by λ^2 and letting $\lambda \to 0$, we find that $x^t H(y) x \geq 0$, which precisely means that H is PSD.

Now suppose that H is PSD $\forall x \in X$. The mean value theorem from calculus allows us to make the following representation:

$$f(x) = f(y) + \nabla f(y)^t(x-y) + \frac{1}{2}(x-y)^t H(z)(x-y) \quad (13)$$

where $z = \lambda x + (1-\lambda)y$ for some $\lambda \in (0,1)$. Since $H(x)$ is PSD and therefore in particular $(x-y)^t H(z)(x-y) \geq 0$, we find that

$$f(x) \geq f(y) + \nabla f(y)^t(x-y). \quad (14)$$

Which by proposition 2.3 is equivalent to convexity for a differentiable function. □

Remark 2.4.1. *If we replace positive definite (PD) with PSD and insert strict inequalities where suitable a similar result holds for PD Hessians implying strict convexity.*

Remark 2.4.2. *If we replace postive semidefiniteness with negative semidefiniteness an equivalent result holds for concavity. This follows immediately by considering $-f$, of which the $Hessian$ matrix then has the opposite sign by linearity of differentiation.*

2.1.2 Duality and Saddle-Point Optimality

Problem 2 (The General Dual Optimization Problem)**.**

$$\begin{aligned} &\max_{u \in \mathbb{R}^m, v \in \mathbb{R}^l} \theta(u, v) \\ &s.t.\ u \geq 0, \end{aligned} \tag{15}$$

where the Langrangian, $\mathcal{L} : \mathbb{R}^{n+m+l} \to \mathbb{R}$ is defined as

$$\mathcal{L}(x, u, v) = f(x) + \sum_{i=1}^{m} u_i g_i + \sum_{j=1}^{l} v_j h_j \tag{16}$$

and the u_i, v_j are the Lagrange mutliplicators of the corresponding g_i, h_j. Using this, the dual function $\theta : \mathbb{R}^{m+l} \to \mathbb{R}$ is defined as

$$\theta(u, v) = \inf_{x \in X} \mathcal{L}(x, u, v). \tag{17}$$

Remark 2.1. *Note that the dual problem is particularly well-behaved in the sense that the problem becomes unconstrained whenever we do not have inequality constraints in the primal problem (1). This follows since if we have no functions g_i, neither do we have corresponding Lagrange multiplicators u_i in the Lagrangian and thus the vectorial constraint $u \geq 0$ is irrelevant.*

We now provide an example which illustrates how to explicitly calculate the dual function.

Example 2.2. *Consider the convex optimization problem to minimize $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 + y^2$ over the convex set $\{(x, y) \in \mathbb{R}^2 : ax^2 + by^2 - 1 \leq 0\}$, with a, b greater than 0. The Lagrangian is*

$$\mathcal{L}(x, y, u) = x^2 + y^2 + u(ax^2 + by^2 - 1). \tag{18}$$

To find the dual function we need to minimize the Lagrangian. Since $u \geq 0, a > 0, b > 0$ we are just minimizing a sum of squares. Clearly $x = y = 0$ and therefore $\theta(u) = -u$.

In order to prove the main results of this section, saddle-point optimality and the KKT-conditions, we first need the following result known as the weak duality theorem.

Theorem 2.5. *Suppose x is feasible to the primal problem and (u, v) is feasible to the dual problem. Then*

$$f(x) \geq \theta(u, v). \tag{19}$$

Proof. The proof is trivial and follows immediately since $\mathcal{L} - f \leq 0$, for all feasible x (i.e. $x \in S$). $\square$

We also have the following useful corollary.

Corollary 2.5.1. *Suppose $(\bar{x}, \bar{u}, \bar{v})$ are feasible to the primal and dual problem respectively and that $f(\bar{x}) = \theta(\bar{u}, \bar{v})$. Then $(\bar{x}, \bar{u}, \bar{v})$ solve the primal and dual problem respectively.*

Proof. $f(x) \geq \theta(u, v)$ for all (x, u, v), but $f(\bar{x}) = \theta(\bar{u}, \bar{v})$. Hence $\bar{x}$ is a global minimum of f. The argument for θ is analogous. $\square$

Definition 2.3. *Given a solution $(\bar{x}, \bar{u}, \bar{v})$, it is called a saddle-point of the Lagrangian $\mathcal{L}$ if $\bar{u} \geq 0$ and*

$$\mathcal{L}(\bar{x}, u, v) \leq \mathcal{L}(\bar{x}, \bar{u}, \bar{v}) \leq \mathcal{L}(x, \bar{u}, \bar{v}) \tag{20}$$

for all $x \in X$, and all (u, v) with $u \geq 0$.

Theorem 2.6. *A solution $(\bar{x}, \bar{u}, \bar{v}), \bar{x} \in X, \bar{u} \geq 0$ is a saddle point of $\mathcal{L}$ if and only if*

- $\mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = \min_{x \in X} \mathcal{L}(x, \bar{u}, \bar{v})$,
- $g(\bar{x}) \leq 0, h(\bar{x}) = 0$ *and*
- $\bar{u}^t g(\bar{x}) = 0$.

Furthermore, $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point if and only if $\bar{x}$ and $(\bar{u}, \bar{v})$ are solutions to the primal and dual problem respectively and without duality gap, that is $f(\bar{x}) = \theta(\bar{u}, \bar{v})$.

Proof. Let $(\bar{x}, \bar{u}, \bar{v})$ be a saddle point of the Lagrangian. Then by the definition above the first condition must be true. Moreover, if (20) is to hold for all $x \in X$, and all (u, v) with $u \geq 0$, we clearly must have that $g(\bar{x}) \leq 0, h(\bar{x}) = 0$ for else we could find (u, v) that violate $\mathcal{L} \leq f$. Using the definition of the Lagrangian, we find by rewriting of (20) that

$$f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) \geq f(\bar{x}) + u^t g(\bar{x}) + v^t h(\bar{x}). \tag{21}$$

Hence the third condition must also hold, for if not then $\bar{u}^t g(\bar{x}) > 0$ for $u = 0$ but by assumption $\bar{u}^t g(\bar{x}) \leq 0$ since $\bar{u} \geq 0, g(\bar{x}) \leq 0$.

Conversely, suppose that $(\bar{x}, \bar{u}, \bar{v})$ fulfills the three conditions for $\bar{x} \in X, \bar{u} \geq 0$. Then $\mathcal{L}(\bar{x}, u, v) \leq f(\bar{x}) = \mathcal{L}(\bar{x}, \bar{u}, \bar{v}) \leq \mathcal{L}(x, \bar{u}, \bar{v})$ shows that $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point of $\mathcal{L}$.

For the second equivalence, suppose again that $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point of $\mathcal{L}$. By assumption and the second property, $\bar{x}$ is feasible to the primal. Since $\bar{u} \geq 0$, $(\bar{u}, \bar{v})$ is feasible to the dual. Further, the first part of the theorem yields that

$$\theta(\bar{u}, \bar{v}) = \mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) = f(\bar{x}). \tag{22}$$

Thus, by the corollary to the weak duality theorem, we know that $(\bar{x}, \bar{u}, \bar{v})$ solve both problems without duality gap.

Last, let $\bar{x}$ and $(\bar{u}, \bar{v})$ be optimal to the primal and dual problem respectively and suppose that $f(\bar{x}) = \theta(\bar{u}, \bar{v})$. Then since $\bar{x}, \bar{u}, \bar{v}$ are part of the optimal solutions, we have that $\bar{x} \in X, g(\bar{x}) \leq 0, h(\bar{x}) = 0$ and $\bar{u} \geq 0$. Now consider

$$\theta(\bar{u}, \bar{v}) = \min_{x \in X} \{f(x) + \bar{u}^t g(x) + \bar{v}^t h(x)\} \leq f(\bar{x}) + \bar{u}^t g(\bar{x}) + \bar{v}^t h(\bar{x}) \leq f(\bar{x}). \tag{23}$$

Since $\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal to the primal and dual respectively, we know that $\bar{u}^t g(\bar{x}) = 0$ and thus, $\mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) = \theta(\bar{u}, \bar{v}) = \min_{x \in X}\{f(x) + \bar{u}^t g(x) + \bar{v}^t h(x)\} = \min_{x \in X} \mathcal{L}(x, \bar{u}, \bar{v})$.

□

Now, it may not always be expedient to search for a saddle point directly. We thus require a method that produces a saddle point given sufficiently nice functions. Reprieve is provided by the Kurosh-Kuhn-Tucker sufficient conditions. This will be our primary workhorse in subsequent optimization problems.

Theorem 2.7. *For differentiable functions f, g and h suppose that $\bar{x} \in S$ and that there exists $\bar{u} \geq 0$ and $\bar{v}$ such that*

$$\begin{aligned} &\nabla \mathcal{L}(\bar{x}, \bar{u}, \bar{v}) = \nabla f(\bar{x}) + \nabla g(\bar{x})^t \bar{u} + \nabla h(\bar{x})^t \bar{v} = 0 \text{ and} \\ &\bar{u}^t g(\bar{x}) = 0. \end{aligned} \tag{24}$$

These are called the KKT-conditions. Assume further that f and g are convex and that h is affine on a nonempty, open convex set X. Then $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point of $\mathcal{L}(\bar{x}, \bar{u}, \bar{v})$.

Proof. Suppose that $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in S$ and $\bar{u} \geq 0$ satisfies the KKT-conditions (24). Then by affinity of h and convexity of f and the g_i we find

$$\begin{aligned} &f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^t (x - \bar{x}) \\ &g_i(x) \geq g_i(\bar{x}) + \nabla g_i(\bar{x})^t (x - \bar{x}) \text{ for } i = 1, .., m \\ &h_i(x) = h_j(\bar{x}) + \nabla h_j(\bar{x})^t (x - \bar{x}) \text{ for } j = 1, .., l \end{aligned} \tag{25}$$

and for all $x \in X$. Multiplying the second and the third equation by $\bar{u}_i$ and $\bar{v}_i$ respectively, and adding them to the first, it follows that $\mathcal{L}(\bar{x}, \bar{u}, \bar{v}) \leq \mathcal{L}(x, \bar{u}, \bar{v})$. Further since $\bar{x} \in S$ we have that $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and by assumption $\bar{u}^t g(\bar{x}) = 0$. It follows that $\mathcal{L}(\bar{x}, u, v) \leq \mathcal{L}(\bar{x}, \bar{u}, \bar{v})$. Hence, $(\bar{x}, \bar{u}, \bar{v})$ a saddle point.

□

The beauty in the above theorem lies in the fact that we have reduced a rather difficult problem, that of finding a saddle point, to one solvable by methods of elementary calculus and linear algebra. I.e. that of finding a point which satisfies the KKT-conditions (24). Such a point is called a *KKT-point*.

2.1.3 Minimizing Concave Functions

Sadly, as we will later show, the fractal dimension of a portfolio is not convex, but concave. This means that we cannot use material just presented to find an optimal solution. However, there is solace in the following theorem.

Definition 2.4. *We say that x is an extreme point of a non-empty convex set, C, if the decomposition*

$$x = \lambda y + (1-\lambda)z \tag{26}$$

for $y, z \in C$, $\lambda \in (0,1)$ implies $x = y = z$.

Theorem 2.8. *Suppose that $\bar{x}$ solves $\min_{x\in C} f(x)$ where $f : C \to \mathbb{R}$ is strictly concave and C is convex and compact. Then $\bar{x}$ is an extreme point of C. In fact, the solution to the problem always lies within the set of extreme points of C.*

Proof. Let $\bar{x}$ be the global minimum and suppose it is not an extreme point. Then we can write $\bar{x} = \lambda y + (1-\lambda)z$ where x, y, z are all distinct and in C. By strict concavity

$$f(\bar{x}) = f(\lambda y + (1-\lambda)z) > \lambda f(y) + (1-\lambda)f(z). \tag{27}$$

Hence, either $f(y)$ or $f(z)$ is less than $f(\bar{x})$ contradicting the fact that $\bar{x}$ solves the problem. We conclude by Weierstrass' theorem that such a solution always exists since C is compact. □

Before we proceed we give an example to illustrate the usefulness of the above theorem.

Example 2.3. *Remember that in particular linear programs have affine, and thus concave, objective functions. It therefore follows from the theorem above that all linear programs of the form $\min_{x\in\mathbb{R}^n} a^t x$ subject to $Ax \le b$ are solved by the extreme points of the set determined by the inequality (we could have equality as well) $Ax \le b$. These ideas are used in for example the simplex method, which is of great practical importance.*

There is further solace in the following proposition. We will show that the set of extreme points of a certain set extremely often considered in portfolio theory is finite and rather trivial.

Proposition 2.9. *Let $C = \{x \in \mathbb{R}^n : x \ge 0, e^t x = 1\}$ where e is a vector of all 1s. Then the set of extreme points of C is given by*

$$C_E = \bigcup_{i=1}^{n} \{x \in C : x_i = 1, x_j = 0 \,\forall j \ne i\}. \tag{28}$$

Proof. Clearly any point of C_E is an extreme point. Now suppose x is an extreme point but $x \notin C_E$, then x can be expressed as the non-negative combination of two points of the hyperplane defined by $e^t x = 1$, but this precisely means that x is not an extreme point. Thus C_E are all the extreme points of C. □

2.2 The Markowitz Model

There is one last stop to be made before starting to consider fractals. Namely Markowitz's original 1952 model [3]. Not only is the Markowitz Model an excellent example of optimization theory at use, it is also the very foundation of our study of optimal portfolios. Our approach and presentation is inspired by [15].

2.2.1 The Basics of Portfolio Theory

Optimal portfolio theory is the mundane study of finding the best, or more technically, optimal, placement of securities for an investor given risk-return preferences using optimization and other mathematics. We again state the problems, basic definitions and assumptions and analyze their necessity.

However, first we wish to mention the *Portfolio Universe*. We define it as the set of securities available in a given optimization problem. It need not be every security available in the market but simply an arbitrary subset. More tangibly, in financial applications it is most often a set of similar assets which are of interest as viewed comparatively, for instance the S&P 500. We now define the portfolio, its associated reward and risk.

Definition 2.5. *We let the portfolio (weighting) be denoted by $x \in \mathbb{R}^n$. Where n is the number of securities in the universe. The return of the portfolio universe is the random vector denoted $r \in \mathbb{R}^n$, and its covariance matrix is $\Sigma \in \mathbb{R}^{n \times n}$, $\Sigma = \mathbb{E}((r - \bar{r})(r - \bar{r})^t)$. Moreover, for convenience, we introduce the vector of all 1's, namely $e \in \mathbb{R}^n$.*

Remark 2.5.1. *Note that we do not require $0 \leq x_i \leq 1$ for $i \in \{1, 2, ..., n\}$ in the definition above. This means that we allow so-called short-selling; the selling of assets that one does not actually own, but borrows.*

Remark 2.5.2. *We try to avoid the word stock in order to allow for more general classes of assets such as risky bonds or oil. However, the essentials are not lost if one just thinks stock, every time one hears asset or security.*

This makes the importance of the following definitions clear.

Definition 2.6. *The reward of a portfolio is the expectation of its return*

$$\rho(x) = \mathbb{E}(r^t x) = \bar{r}^t x. \tag{29}$$

Definition 2.7. *The risk of a portfolio is the variance of its return*

$$R(x) = \mathrm{Var}(r^t x) = \mathbb{E}((r^t x - \mathrm{E}(r^t x))^2) = \mathbb{E}(x^t (r - \bar{r})(r - \bar{r})^t x) = x^t \Sigma x \tag{30}$$

We are now ready to present the first and simplest problem of portfolio theory with n securities. However, before we start, for the sake of completeness, we digress shortly on estimation.

2.2.2 Estimation of Mean and Variance

We do not make any explicit assumption about the origin of the return vector. For our purposes it suffices to assume that it some known probability distribution with finite variance such that its expectation and covariance matrix are available to us. Moreover, this thesis is predominantly theoretical and does not treat analysis of data in detail. Nevertheless, Portfolio Theory is an inherently applied subject and its practical uses depend heavily on the estimation of the return vector. In general, one uses a financial time series to estimate returns by historical averages. These are then, in some sense, projected into the future as an estimate for the expected return.

A mathematical description of a time series as the graph of a function may be found in section 3. For now, we consider a time series as a collection of ordered measurement points for some quantity between time 1 and T, all integers.

Definition 2.8. *The sample expected return vector is given by*

$$\bar{r} = \frac{1}{T}\sum_{i=1}^{T} r(i). \tag{31}$$

Remark 2.8.1. *We use the notation $r(i)$ instead of r_i since the subscript may be confused with meaning the return of the i^{th} asset.*

Definition 2.9. *The sample covariance matrix is given by*

$$\Sigma = \frac{1}{T-1}\sum_{i=1}^{T}(r(i)-\bar{r})^t(r(i)-\bar{r}). \tag{32}$$

2.2.3 Discussion and Solution of the n-security problem

In the problem, preferences are such that there exists a linear preference for portfolio return whereas disfavor of risk is quadratic. Moreover, we use a scalar trade-off factor, μ, for return which captures the inherent relative preferences for risk and return of the person and/or firm facing the optimization problem. Lastly, we normalize the budget constraint to 1. Each entry of the optimal portfolio $\bar{x}$ can then be interpreted as the fraction of available resources to be invested in a certain security.

Problem 3 (The n-Security Problem)**.**

$$\begin{aligned} &\min_x \frac{R(x)}{2} - \mu\rho(x) \\ &s.t.\ e^t x = 1 \end{aligned} \tag{33}$$

In order for our problems to be well-behaved we need further assumptions. The first of which is a slight restriction on the covariance matrix. Since $\sigma(r_i, r_j) = \sigma(r_j, r_i)$ and since for arbitrary $x \in \mathbb{R}^n$

$$x^t\Sigma x = x^t(r-\bar{r})(r-\bar{r})^t x = (x^t(r-\bar{r}))^2 \geq 0 \tag{34}$$

we have that Σ is symmetric positive semidefinite. We further impose the restriction that all assets are in fact risky. This corresponds to:

Assumption 1. *The covariance matrix is positive definite. That is,* $\Sigma > 0$.

The absence of risk is in our context precisely means zero variance (of at least one asset), this means that the covariance matrix of the universe has only non-zero eigenvalues under the assumption of riskiness. In particular Σ has rank n so that assumption 1 implies the existence of an inverse Σ^{-1}. Moreover, semi-definiteness already guarantees that our class of problems is convex. In particular, they are solved analytically without excessive trouble since they are also differentiable. This brings us to

Lemma 2.10. *Any problem with objective function of the form*

$$\pi(x) = x^t Ax + b^t x + c \tag{35}$$

where $A \in \mathbb{S}^n_+$ *(the set of symmetric positive semi-definite matrices),* $b \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$ *is convex and differentiable.*

Proof. Differentiability follows immediately from differentiability of polynomials. Hence, we can find the Hessian of $\pi(x)$ as

$$H_\pi(x) = J(\nabla\pi(x)) = J(\nabla(x^t Ax + b^t x + c)) = J(2Ax - b) = J(2Ax) = 2A$$

which was assumed to be in $\mathbb{S}^n_+$ and hence convexity follows from proposition 2.4. $J(\cdot)$ denotes the Jacobian matrix. □

Assumption 2. *The return is not a multiple of* $e = (1, ...1)^t$*. I.e.* $\bar{r} \neq ce, \forall c \in \mathbb{R}$.

Proposition 2.11. *Suppose assumption 1 holds. Then (33) has the unique primal and dual solution*

$$\bar{x} = \Sigma^{-1}(\bar{\lambda}e + \mu\bar{r}), \quad \bar{\lambda} = \frac{1 - \mu e^t \Sigma^{-1}\bar{r}}{e^t \Sigma^{-1} e} \tag{36}$$

and associated return

$$\rho(x) = \bar{r}^t \Sigma^{-1}(\bar{\lambda}e + \mu\bar{r}). \tag{37}$$

Proof. We will use saddle-point optimality. The Lagrangian of (33) is

$$\mathcal{L}(x, \lambda) = \frac{x^t \Sigma x}{2} - \mu\bar{r}^t x - \lambda(e^t x - 1) \tag{38}$$

and has gradient

$$\nabla\mathcal{L}(x, \lambda) = \nabla\frac{x^t \Sigma x}{2} - \nabla\mu\bar{r}^t x - \nabla\lambda(e^t x - 1) = \Sigma x - \mu\bar{r} - \lambda e. \tag{39}$$

Setting $\nabla\mathcal{L}(x, \lambda) = 0$ (dual feasibility) and applying left multiplication of Σ^{-1} yields

$$x = \Sigma^{-1}(\lambda e + \mu\bar{r}). \tag{40}$$

Substituting $\bar{x}$ into the budget constraint yields the primal feasiblility equation

$$\begin{aligned} &e^t(\Sigma^{-1}(\lambda e + \mu\bar{r})) - 1 = 0 \\ \Leftrightarrow &e^t(\Sigma^{-1}(\lambda e + \mu\bar{r})) - \frac{e^t\Sigma^{-1}e}{e^t\Sigma^{-1}e} = e^t\Sigma^{-1}\left(\lambda e + \mu\bar{r} - \frac{e}{e^t\Sigma^{-1}e}\right) = 0 \end{aligned} \tag{41}$$

We now try to find a root by considering only the parenthesis. We have that

$$\lambda e = \frac{e}{e^t\Sigma^{-1}e} - \mu\bar{r}. \tag{42}$$

Left multiplication by $\frac{e^t\Sigma^{-1}}{e^t\Sigma^{-1}e}$ gives

$$\bar{\lambda} = \frac{1 - \mu e^t\Sigma^{-1}\bar{r}}{e^t\Sigma^{-1}e}. \tag{43}$$

Further, we note that substituting $\lambda = \bar{\lambda}$ solves the original equation. We now finish by noting that since the problem (33) is convex by lemma 2.10, and since $(\bar{x}, \bar{\lambda})$ is a KKT point the main result follows by theorem 2.7. Hence, $(\bar{x}, \bar{\lambda})$ is a saddle point and the unique optimal solution to the primal and dual problem without duality gap. The associated return follows by substition of $\bar{x}$.

□

2.2.4 Introducing Risk-Free Assets

We will now consider problem 4, which essentially is (33) with cash. Cash is typically some fixed-income security such as government T-bills which is regarded as risk-free. Technically, nothing is truly risk-free however the default risk of most industrialized nations is sufficiently small such that the default risk is negligible and the approximation sensible. We define the new portfolio $z = (x, x^c)$ where x^c is the amount invested (or borrowed) in cash. Moreover, set $\bar{a} = (\bar{r}, r^c)$. We let $r^c = \bar{r}^c$ be the return of cash and $\Sigma_z = \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix}$ be the associated covariance matrix. Hence, it is obvious that cash does not affect portfolio variance directly, i.e.

$$R(z) = z^t \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} z = x^t\Sigma x = R(x). \tag{44}$$

We now formulate and solve problem 4.

Problem 4.

$$\begin{aligned} &\min_z \frac{z^t\Sigma_z z}{2} - \mu\bar{a}^t z \\ &s.t.\ e^t z = 1 \end{aligned} \tag{45}$$

Proposition 2.12. *Suppose assumption 1 holds. Then (45) has the unique primal and dual solution*

$$\bar{z} = (\bar{x}^t, x^c)^t = \left(\mu\Sigma^{-1}(\bar{r} - er^c), 1 - \mu e^t\Sigma^{-1}(\bar{r} - er^c)\right), \quad \bar{\lambda} = -\mu r^c \tag{46}$$

Proof. The Lagrangian is

$$\begin{aligned}\mathcal{L}(z,\lambda) &= \frac{z^t \Sigma_z z}{2} - \mu \bar{a}^t z - \lambda(e^t z - 1) \\ &= \frac{x^t \Sigma x}{2} - \mu(\bar{r}^t, r^c)(x^t, x^c)^t - \lambda(e^t x + x^c - 1)\end{aligned} \tag{47}$$

which has gradient

$$\nabla \mathcal{L}(z,\lambda) = \Sigma_z z - \mu \bar{a} - \lambda e = \Sigma_z (x^t, 0)^t - \mu(\bar{r}^t, r^c)^t - \lambda e. \tag{48}$$

Applying dual feasibility, $\nabla \mathcal{L}(z,\lambda) = 0$, and multiplying from the left by

$$\begin{pmatrix} \Sigma^{-1} & 0 \\ 0 & 1 \end{pmatrix} \tag{49}$$

yields

$$x = \Sigma^{-1}(\mu \bar{r} + \lambda e) \tag{50}$$

and by considering the last row of the gradient of the Lagrangian, we get

$$\mu r^c + \lambda = 0 \Leftrightarrow \lambda = -\mu r^c. \tag{51}$$

Substituting this x into the budget constraint gives us quite naturally that the difference between a full portfolio and that which was invested in risky assets is, in fact, invested (borrowed) in cash

$$e^t(\Sigma^{-1}(\mu \bar{r} + \lambda e), x^c) = 1 \Leftrightarrow x^c = 1 - e^t \Sigma^{-1}(\mu \bar{r} + \lambda e). \tag{52}$$

To finish the proof, note that we have now shown that $(\bar{z}, \lambda)$ satisfy the KKT conditions. By lemma 2.10 the problem is convex. Now note that the assumptions of theorem 2.7 are fulfilled and $(\bar{z}, \bar{\lambda})$ is thus a saddle-point and the unique optimal solution to the primal and dual problem respectively without duality gap.

□

Remark 2.12.1. *For $r^c = 0$ this returns the risky Markowitz solution. This also makes intuitive sense. There is no reason to invest in an asset if it neither provides positive return, nor any hedging (remember that the last row and column of the covariance matrix is identically* 0*).*

We wish to comment further on the assumption of cash as risk-free. If we instead consider cash a risky asset in problem (33) we would face the problem that the risk would be very low in most cases and thus the eigenvalue of Σ corresponding to cash would be very small. Equivalently, the cash eigenvalue of Σ^{-1} would be very large and hence since the placement in cash is proportional to the cash eigenvalue of Σ^{-1} would be very sensitive to small errors in the practical approximation of the riskiness of cash. The approximation thus also makes sense from a practical standpoint as to minimize the error bars of the portfolio variance.

2.3 Fractals and Fractal Dimensions

Before proceeding with the main topic, we need to introduce the concepts of fractal and fractal dimension. We introduce merely the basics. Our account is based on [16] and further knowledge is found within.

The fractal dimension is often represented as the Box-Counting dimension. We will prefer its use over the Hausdorff-Besicovitch dimension due to the geometric intuition of the previous. For many simple cases, including fractional Brownian motion, as defined below, they are equal. Even though we will primarily be counting boxes, the other definition is also of interest as some theoretical aspects are better treated.

Definition 2.10. *Let F be a bounded non-empty subset of $\mathbb{R}^n$ and let $N_\delta(F)$, $\delta > 0$ be the smallest number of n-dimensional boxes of sides of length at most δ which can cover F. The upper and lower Box-Counting dimensions of F, are defined as*

$$\overline{D_B}(F) = \limsup_{\delta \to 0^+} \frac{\ln N_\delta(F)}{-\ln \delta}, \tag{53}$$

and

$$\underline{D_B}(F) = \liminf_{\delta \to 0^+} \frac{\ln N_\delta(F)}{-\ln \delta}. \tag{54}$$

Provided both limits exist and are equal, the Box-Counting dimension is then

$$D_B(F) = \lim_{\delta \to 0^+} \frac{\ln N_\delta(F)}{-\ln \delta}. \tag{55}$$

We wish to establish some basic machinery required for performing our dimensional calculations. As such, the Hausdorff-Besicovitch Dimension, the related measure, and some elementary properties are presented below.

Definition 2.11. *Let $F \subset \mathbb{R}^n$ and $s \geq 0$. Then for any $\delta > 0$ we define*

$$\mathcal{H}^s_\delta(F) = \inf\left\{\sum_{i=1}^{\infty} (\text{diam } U_i)^s : \{U_i\} \text{ is a } \delta\text{-cover of } F\right\}. \tag{56}$$

Then we denote $\mathcal{H}^s(F) = \lim_{\delta\to 0} \mathcal{H}^s_\delta(F)$ the s-dimensional Hausdorff measure of F.

We can then define the Hausdorff-Besicovitch Dimension, D_H, as

$$D_H(F) = \inf\{s \geq 0 : \mathcal{H}^s(F) = 0\}. \tag{57}$$

We remind the reader that for a metric space X, a subset F of X has diameter defined as the supremum over all distances between elements of F.

Remark 2.11.1. *For any set F we have that $D_H(F) \leq D_B(F)$ since the boxes of D_B also constitute a δ-covering and we take the infimum over all such coverings when constructing D_H.*

Even though our study concerns the generalized notion of self-affine processes, it can still be useful to have an intuitive notion of what is a fractal. We will not give a mathematical definition as they often are different, for different authors. For example, one can think of a fractal subset of Euclidean space $\mathbb{R}^2$ to be one in which the fractal dimension $D_H \in (1,2)$ and if we "look" at a small part of the set, it should look like a scaled copy of the whole. However, this should only be regarded as an intuitive image and not a formal definition.

We now give two examples.

Example 2.4 (The Cantor Set). *Consider the interval* $[0,1]$. *If we cut away the middle third of length* $1/3$ *we are left with two bars of length* $1/3$ *each call them* c_1, c_2. *If we repeat this process for* c_1, c_2 *and so on, ad infinitum, we obtain the cantor set. Interestingly, the dimension of this set is positive, and not zero as one may expect from a union of essentially point-like sets. Mandelbrot illustrates in [2] that is, in fact,* $\frac{\ln 2}{\ln 3}$.

Example 2.5 (The Mandelbrot Set). *Let* $P_c : \mathbb{C} \to \mathbb{C}$ *be defined by* $P_c(z) = z^2 + c$ *and let* $P_c^n = P_c \circ ... \circ P_c$ *(i.e. composed with itself* n *times). The Mandelbrot set is then defined as*

$$\mathbb{M} = \{c \in C : \exists s \in \mathbb{R}, \forall n \in \mathbb{N}, |P_c^n(0)| \leq s\}. \tag{58}$$

The Mandelbrot set also emphasizes the point that what we called a fractal above is only the intuition. It is shown in [18] that the boundary of the Mandelbrot set has Hausdorff dimension equal to 2. *For a visualization generated in MATLAB we refer to figure 1.*

2.3.1 Techniques for Calculating Dimensions

Certain mappings under stronger continuity assumptions do not alter the dimension of a set. In general we are interested in those that are Hölder or Lipschitz continuous .

Definition 2.12. *Let* $F \subset \mathbb{R}^n$ *and* $f : F \to \mathbb{R}^m$. *The function* f *is said to be Hölder continuous of order* α *on* F *if there exist constants* $c > 0, \alpha > 0$ *such that*

$$|f(x) - f(y)| \leq c|x - y|^\alpha \tag{59}$$

for all $x, y \in F$. *If the relation holds for* $\alpha = 1$ *the mapping is said to be Lipschitz. Moreover, if for some* $c > 0$

$$\frac{1}{c}|x - y| \leq |f(x) - f(y)| \leq c|x - y| \tag{60}$$

for all $x, y \in F$ *the function is said to be (c-)bi-Lipschitz on* F.

Lemma 2.13. *Let* $F \subset \mathbb{R}^n$ *and* $f : F \to \mathbb{R}^m$ *be a mapping such that* f *fulfills a Hölder-condition of order* α *(for* $\alpha = 1$ *this is ordinary Lipschitz continuity). Then* $\forall s \geq 0$

$$\mathcal{H}^{s/\alpha}(f(F)) \leq c^{s/\alpha}\mathcal{H}^s(F). \tag{61}$$

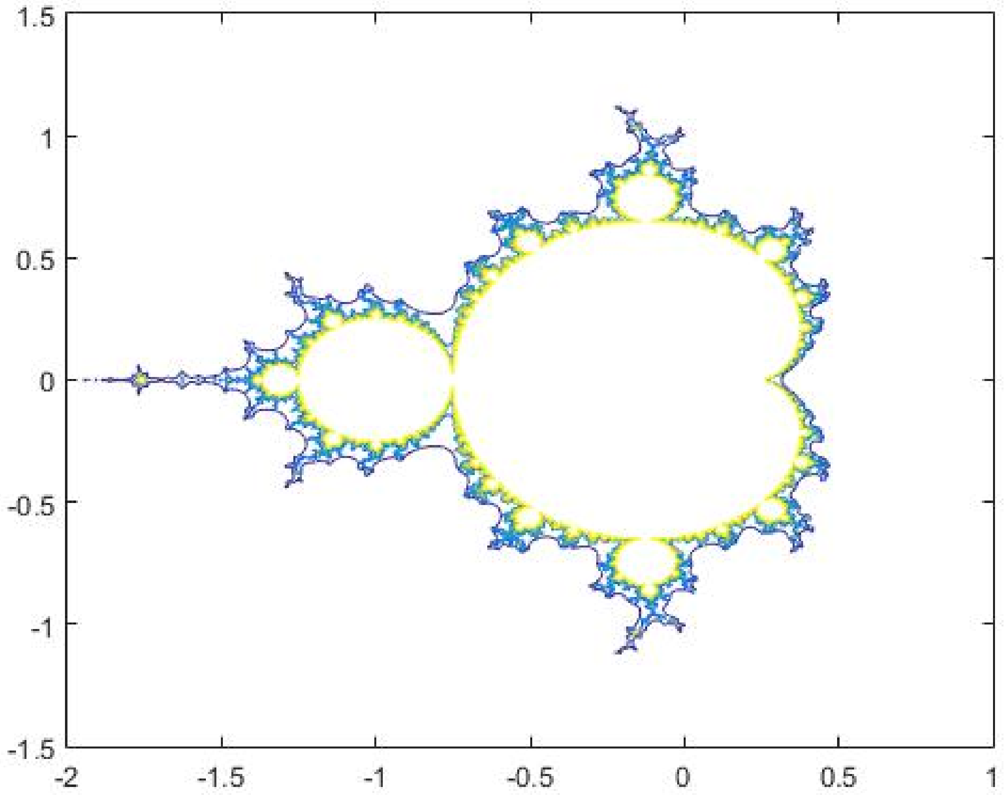

Figure 1: The Mandelbrot set. Notice how each small "blob" resembles the whole. This is at the very essence of fractal geometry. MATLAB code for how to generate the image above can be found in chapter 13 of [17].

Proof. As $\{U_i\}$ is a δ-cover of F and since

$$\text{diam } f(F \cap U_i) \leq c(\text{diam } F \cap U_i)^\alpha \leq c(\text{diam } U_i)^\alpha, \tag{62}$$

it follows that $\{f(F \cap U_i)\}$ is an $c\delta^\alpha$-cover of $f(F)$. Hence,

$$\sum_i (\text{diam } f(F \cap U_i))^{s/\alpha} \leq c^{s/\alpha} \sum_i (\text{diam } U_i)^s. \tag{63}$$

Therefore, $\mathcal{H}^{s/\alpha}_{c\delta^\alpha}(f(F)) \leq c^{s/\alpha} \mathcal{H}^s_\delta(F)$. The result follows by taking $\delta \to 0$.

□

Lemma 2.14. *Let $F \subset \mathbb{R}^n$ and suppose that $f : F \to \mathbb{R}^m$ satisfies a Hölder condition of order α. Then $D_H f(F) \leq (1/\alpha) D_H(F)$.*

Proof. For $s > D_H(F)$ the previous lemma yields $\mathcal{H}^{s/\alpha}(F) \le c^{s/\alpha}\mathcal{H}^s(F) = 0$. Hence by definition $D_H(f(F)) \le s/\alpha$. Certainly, $s \le D_H(F)$ and the result follows.

□

Theorem 2.15. *Let F be a bounded set subset of X of fractal dimension D and suppose f is bi-Lipschitz on F. Then the image $f(F)$ also has fractal dimension D.*

Proof. Take $\alpha = 1$ (and $c = 1$) in the lemma above and apply to $f^{-1} : f(F) \to F$. □

Note in particular that all affine (and thus all linear) transformations are covered by the above proposition. We can think of this as meaning that we can stretch and shift graphs at will without affecting the dimension to be calculated.

As financial time series can from our viewpoint be interpreted as graphs of continuous functions, the dimensional properties of the latter are of obvious interest.

Proposition 2.16. *Let $f : [0,1] \to \mathbb{R}$ be continuous and let $\delta \in (0,1)$. We define m to be the least integer greater than or equal to $1/\delta$. Then if N_δ is the number of squares of the δ-mesh that interesect the graph of f,*

$$\delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta] \le N_\delta \le 2m + \delta^{-1} \sum_{i=0}^{m-1} R_f[i\delta, (i+1)\delta] \tag{64}$$

where

$$R_f[t_1, t_2] = \sup_{t_1 \le t, u \le t_2} |f(t) - f(u)| \tag{65}$$

is the maximum range over the interval $[t_1, t_2]$.

Proof. Consider an interval of length δ. The number of of δ-meshes that intersects f on such on interval is between $R_f[i\delta, (i+1)\delta]\delta^{-1}$ and $R_f[i\delta, (i+1)\delta]\delta^{-1} + 2$. The proposition follows by summing over all such intervals. □

The immediate application of the above proposition is to find an upper bound for the dimension of the graph of a function f.

Corollary 2.16.1. *Let $f : [0,1] : \mathbb{R}$ and suppose that f fulfills a Hölder condition of $2 - s$, $1 \le s \le 2$. Then the graph of f has Hausdorff-Besicovitch and Box-Counting dimensions less than or equal to s.*

Proof. By definition we have that $R_f[t_1, t_2] \le c|t_1 - t_2|^{2-s}$ for some $c \in R$. Then by (64)

$$N_\delta \le 2m + \delta^{-1} m c \delta^{2-s} \le (1 + \delta^{-1})(2 + c\delta^{-1}\delta^{2-s}) = (1 + \delta^{s-1})(2 + c\delta)\delta^{-s} \le e\delta^{-s} \tag{66}$$

for sufficiently small $\delta > 0$. Then

$$\frac{\ln N_\delta}{-\ln \delta} \le s \tag{67}$$

and the result follows from the definition 2.10 and the fact that $D_B \ge D_H$. □

2.4 Joseph and Noah effects

The Joseph and Noah effects are allusions to biblical stories [6]. The naming stems from the old testament in which Joseph was a slave who prophecized that there would be seven years of prosperity following seven years of famine; it is thus used to describe anti-persistent behavior in financial markets. Similarly, the Noah effect stems from the story of Noah and his ark. Allegedly, God flooded the earth in Noah's sixth hundred year. In finance, it is used to describe market crashes, which lie in the tails of return distributions.

We now extend the ideas of fractal geometry from the previous section to random variables. Mandelbrot et al. ([19], [20]) define self-affine processes as in the following.

Definition 2.13. *Given $X(0) = 0$, a stochastic process is called self-affine if there exists $\alpha > 0$ such that*

$$\{X(ct_1), ..., X(ct_k)\} \stackrel{d}{=} \{c^\alpha X(t_1), ..., c^\alpha X(t_k)\} \tag{68}$$

for all $c, t_i \geq 0$ where $i = 1, 2, 3, ..., k$. We call α the (self-)affinity index.

Remark 2.13.1. *Some authors, especially within probability refer to such processes as self-similar (for instance those in the bibliography). We have chosen to use Mandelbrot's convention since our work is primarily based on his.*

By no means we wish to suggest that there is an actual stochastic process driving the stock market, or at least not one that we can know. Nevertheless, to illustrate the Joseph and Noah effects we need a mathematical model of market returns. Moreover, B. B. Mandelbrot demonstrates in [6] that certain self-affine processes bare high resemblance to the actual behavior of the market.

2.4.1 Fractional Brownian Motion and the Joseph Effect

Definition 2.14. *Fractional Brownian motion (fBm) of index α ($0 < \alpha < 1$) is a stochastic process $X : [0, \infty) \to \mathbb{R}$ such that:*

- *It holds almost surely that $X(t)$ is continuous and that $X(0) = 0$.*
- *$\forall t \geq 0$ and $h > 0$, the increments, $X(t + h) - X(t)$, are stationary and normally distributed with mean 0 and variance $h^{2\alpha}$, such that*

$$P(X(t + h) - X(t) \leq x) = \frac{h^{-\alpha}}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left((-y^2/2h^{2\alpha})\right) dy. \tag{69}$$

Remark 2.14.1. *Note that usual Brownian motion is characterized by $\alpha = 1/2$ and thus this definition is a generalization of standard Brownian motion.*

Proposition 2.17. *Fractional Brownian motion of index α is self-affine with affinity index α.*

Proof. The mean and variance of a Gaussian process uniquely determine the process, see for instance [21]. From definition 2.14 it is clear that both $\{X(ct)\}$ and $\{c^\alpha X(t)\}$ have the same mean and variance. Hence they are equal in distribution. $\square$

Proposition 2.18. *Fractional Brownian motion of order α satisfies a Hölder condition of order α almost surely.*

Proof. Let $0 < h < t$ and $0 < \gamma < \alpha$. By self-affinity and stationary increments

$$\mathbb{E}[|X(t) - X(h)|^{1/\gamma}] = \mathbb{E}[|X(t-h)|^{1/\gamma}] = |t-h|^{\alpha/\gamma}\mathbb{E}[|X(1)|^{1/\gamma}]. \tag{70}$$

The result follows by application of the Kolmogorov-Centsov Theorem as found in [22] page 53. $\square$

Proposition 2.19. *The Box-Counting dimension D_B, and the Hausdorff dimension, of the graph of fractional Brownian motion is almost surely $2-\alpha$.*

Proof. We acquire an upper bound by corollary 2.16.1 and the fact that index-α fBm satisfies a Hölder condition of order α almost surely.

We will construct a mass distribution on the graph with finite s-energy for $s < 2-\alpha$ to obtain a lower bound on the fractal dimension.

If we define $r = |X(t+h) - X(t)|$ and $r = w^{1-\alpha}h^\alpha$ we get the estimate

$$\begin{aligned}
&\mathbb{E}\left[(|X(t+h) - X(t)|^2 + h^2)^{-s/2}\right] \\
=&\int_0^\infty (r^2+h^2)^{-s/2}\frac{h^{-\alpha}}{\sqrt{2\pi}}\exp\left(\frac{-r^2}{2h^{2\alpha}}\right)dr \\
=&\frac{h^{-\alpha}}{\sqrt{2\pi}}\int_0^\infty (r^2+h^2)^{-s/2}\exp\left(\frac{-r^2}{2h^{2\alpha}}\right)dr \\
=&\frac{h^{-\alpha}}{\sqrt{2\pi}}\int_0^\infty (w^{2-2\alpha}h^{2\alpha}+h^2)^{-s/2}\exp\left(\frac{-w^{2\alpha}}{2}\right)\frac{w^{-\alpha}h^\alpha}{2}dw \\
=&\frac{1}{2\sqrt{2\pi}}\int_0^\infty (w^{2-2\alpha}h^{2\alpha}+h^2)^{-s/2}\exp\left(\frac{-w^{2\alpha}}{2}\right)w^{-\alpha}dw \\
\le& c\int_0^h (h^2)^{-s/2}w^{-\alpha}dw + c\int_h^\infty (w^{2-2\alpha}h^{2\alpha})^{-s/2}w^{-\alpha}dw \\
\le& c_1 h^{1-\alpha-s}
\end{aligned} \tag{71}$$

for some $c, c_1 > 0$. Now, it is easy to check that $\mu_f(A) = L\{t \in [0,1] : (f, f(t)) \in A\}$ is a mass distribution on the graph of f in accordance with definition A.2. Then the associated s-energy (for definition see the appendix) is

$$
\begin{aligned}
I_s &= \mathbb{E}\left(\int\int |x-y|^{-s} d\mu_X(x) d\mu_Y(y)\right) \\
&= \int_0^1 \int_0^1 \mathbb{E}\left[(|(X(t)-X(u)|^2 + |t-u|^2)^{s/2}\right] dtdu \\
&\leq \int_0^1 \int_0^1 c_1 |t-u|^{1-\alpha-s} dtdu < \infty
\end{aligned}
\tag{72}
$$

whenever $s < 2-\alpha$. It then follows by proposition A.2 that $D_H(\text{graph } f) \geq 2-\alpha$. Hence $D_H(\text{graph } f) = 2-\alpha$ almost surely.

□

Now we have shown what we set out to do, namely that we have shown that the index α specified in the distribution of the increments of fBm is precisely its fractal dimension. Next, we wish to tie this rather theoretical aspect of the process (69) to the return movements of financial data. To do this, we first define the autocorrelation and autocovariance functions.

Definition 2.15. *The Autocovariance function of a stochastic process, $X(t)$, is given by*

$$R(t_1, t_0) = Cov(X(t_1), X(t_0)). \tag{73}$$

The Autocorrelation function of a Stochastic Process $X(t)$ is given by

$$\rho(t_1, t_0) = \frac{Cov(X(t_1), X(t_0))}{\sqrt{Var(X(t_1))Var(X(t_0))}}. \tag{74}$$

When $t_0 = 0$ we may drop the second variable for convenience.

Proposition 2.20. *The autocorrelation function of index-α fBm is given by*

$$\rho(t) = \frac{1}{2}\left((t+1)^{2\alpha} - 2t^{2\alpha} + (t-1)^{2\alpha}\right). \tag{75}$$

Proof. The autocovariance function for the increments, $\Delta X(t)$, of fBm is defined as

$$
\begin{aligned}
\rho(t) &= \frac{\mathbb{E}[((X(t+1)-X(t))(X(1)-X(0))]}{\sqrt{\mathbb{E}[(X(t+1)-X(1))^2]\mathbb{E}[(X(1)-X(0))^2]}} \\
&= \frac{\mathbb{E}[((X(t+1)-X(t))X(1)]}{E[X(1)^2]} = \frac{\mathbb{E}[(X(t+1)X(1)-X(t)X(1))]}{E[X(1)^2]} \\
&= \frac{\left((t+1)^{2\alpha} - 2t^{2\alpha} + (t-1)^{2\alpha}\right)\mathbb{E}[X(1)^2]}{2\mathbb{E}[X(1)^2]}
\end{aligned}
\tag{76}
$$

using that $X(0) = 0$ almost surely, that $\mathbb{E}[X(t)] = 0$, and the fact that the process is self-affine. □

Corollary 2.20.1. *Note in particular that for $\alpha \in [0,1]$,*

$$\begin{aligned} \alpha > 1/2 \quad &\Rightarrow \rho > 0, \\ \alpha = 1/2 \quad &\Rightarrow \rho = 0, \\ \alpha < 1/2 \quad &\Rightarrow \rho < 0. \end{aligned} \tag{77}$$

Moreover, $\sum_{t=1}^{\infty} |\rho(t)|$ is finite if and only if $\alpha \leq 1/2$.

Remark 2.20.1. *The property $\sum_{t=1}^{\infty} |\rho(t)| = \infty$ is referred to as long-range dependence. The intuition is that if $\lim_{t\to\infty} |\rho(t)| = 0$ sufficiently slow, this means not only that the sum does not converge but also that there is correlation between increments separated very far by time.*

Remark 2.20.2. *$\alpha = 1/2$ yields the expected result of uncorrelated increments, as exhibited by standard Brownian motion.*

From the corollary above in combination with proposition 2.19 we immediately see that for a fractal Brownian motion $\{X\}_{t\geq 0}$, $D_H(\text{graph } X) > 1.5$ means that the process is anti-persistent; it exhibits the Joseph effect. Similiarily, for $D_H(\text{graph } X) < 1.5$ the process is persistent, and does in fact exhibit long-range dependence.

Example 2.6. *Consider the simple stock price model $S(t) = e^{rtX(t)}$ where $X(t)$ is fractional Brownian motion of index α and the return parameter $r > 0$. Together, the multiple $rX(t)$ can be thought of to be the time-varying return. Then*

$$\mathbb{E}[S(t)] = S(0)\int_0^{\infty} \exp(rtx)\exp\left(\frac{-x^2}{2t^{2\alpha}}\right)dx = S(0)\int_0^{\infty} \exp\left(rtx - \frac{x^2}{2t^{2\alpha}}\right)dx. \tag{78}$$

Note that the integral on the right hand is increasing in α whenever $t \geq 1$. This reflects the fact that we should expect a much more "undisturbed" growth of the stock price whenever we have persistent behavior and long-range dependence.

2.4.2 Lévy Processes and the Noah Effect

Just a fBm is a natural generalization of the normal random walk to include persistent and anti-persistent behavior (correlated increments), we can also instead generalize to Lévy Processes to include fat tails. We also show the Noah effect, fat tails, is characterized by the Hausdorff dimension of the graph of Lévy stable processes.

Definition 2.16. *A Lévy process is a stochastic process, $\{X(t)\}_{t\geq 0}$, that satisfies:*

- *It holds almost surely that $X(0) = 0$ and that $X(t)$ is continuous; that is for all $\epsilon > 0, \lim_{s\to t} P(|X(s) - X(t)| > \epsilon) = 0$.*
- *The increments $X(t+\tau) - X(t)$ are independent of t for all $t, \tau \geq 0$. We can write $X(t+\tau) - X(t) = X(\tau)$.*
- *The increments $X(t+\tau) - X(t)$ and $X(s+\zeta) - X(s)$ are independent for all $s, t, \tau, \zeta \geq 0$, $t \neq s$.*

Such a process is called stable if it also adheres to the definition below. We apologize in advance for the notational overload on the parameter α and ask the reader to note that its use in this section differs from the previous. We use the two conventions as they are standard in their respective (sub-)fields. In fact, the two uses obey an inverse relation.

Definition 2.17. *A random variable (or vector) X has a stable distribution if there exist $a, b, c \in \mathbb{R}_{>0}$ and $c_1 \in \mathbb{R}$ such that*

$$aX_1 + bX_2 \stackrel{d}{=} cX + c_1 \tag{79}$$

where X_1, X_2 are independent copies of X.

A random variable is called stable of index α or simply α-stable if the number a, b, c satisfy

$$a^\alpha + b^\alpha = c^\alpha \tag{80}$$

For $\alpha \in (0, 2]$. Further, this generalizes naturally to stochastic processes, which we call stable if the random vectors $(X(t_1), ..., X(t_d))$ are stable for all $t_1 < ... < t_d$.

Example 2.7 (Normal Distribution)**.** *Consider three independent and normally distributed random variables X, X_1, X_2 all with mean μ and variance σ^2. Then $aX_1 + bX_2$ is normally distributed with mean $(a + b)\mu$ and variance $(a^2 + b^2)\sigma^2$. Therefore, $aX_1 + bX_2 \stackrel{d}{=} (a^2 + b^2)^{1/2}X + (a + b - (a^2 + b^2)^{1/2})\mu$. Comparing with definition 2.17 we note that the normal distribution is α-stable with $\alpha = 2$.*

A plot of the probability density functions of the normal ($S_1(1, 0, 0)$) and Cauchy ($S_2(1, 0, 0)$) distributions can be found in figure 2.

We now introduce characteristic functions to give another alternate approach to stable distributions.

Definition 2.18. *The characteristic function of a random variable X is given by*

$$\phi_X(u) = \mathbb{E}[e^{iuX}], u \in \mathbb{R}. \tag{81}$$

Definition 2.19 (equivalent to definition 2.17)**.** *A random variable X is said to have a stable distribution if there are parameters $\alpha \in (0, 2], \sigma \geq 0, \beta \in [-1, 1]$ and $\mu \in \mathbb{R}$ such that the characteristic function has following form:*

$$\phi_X(\theta) = \mathbb{E}[e^{i\theta X}] = \begin{cases} \exp\left(-\sigma^\alpha|\theta|^\alpha(1 - i\beta\mathrm{sgn}(\theta)\tan(\frac{\pi\alpha}{2}) + i\mu\theta)\right) & \textit{if } \alpha \neq 1 \\ \exp\left(-\sigma^\alpha|\theta|^\alpha(1 + i\beta\mathrm{sgn}(\theta)(\frac{2}{\pi})\ln|\theta| + i\mu\theta)\right) & \textit{if } \alpha = 1 \end{cases} \tag{82}$$

Remark 2.19.1. *Since α and (σ, β, μ) completely characterize stable random variables we can for an apporiate stable X write $X \sim S_\alpha(\sigma, \beta, \mu)$.*

Remark 2.19.2. *The definitions 2.17 and 2.19 are equivalent. More details can be found in [23].*

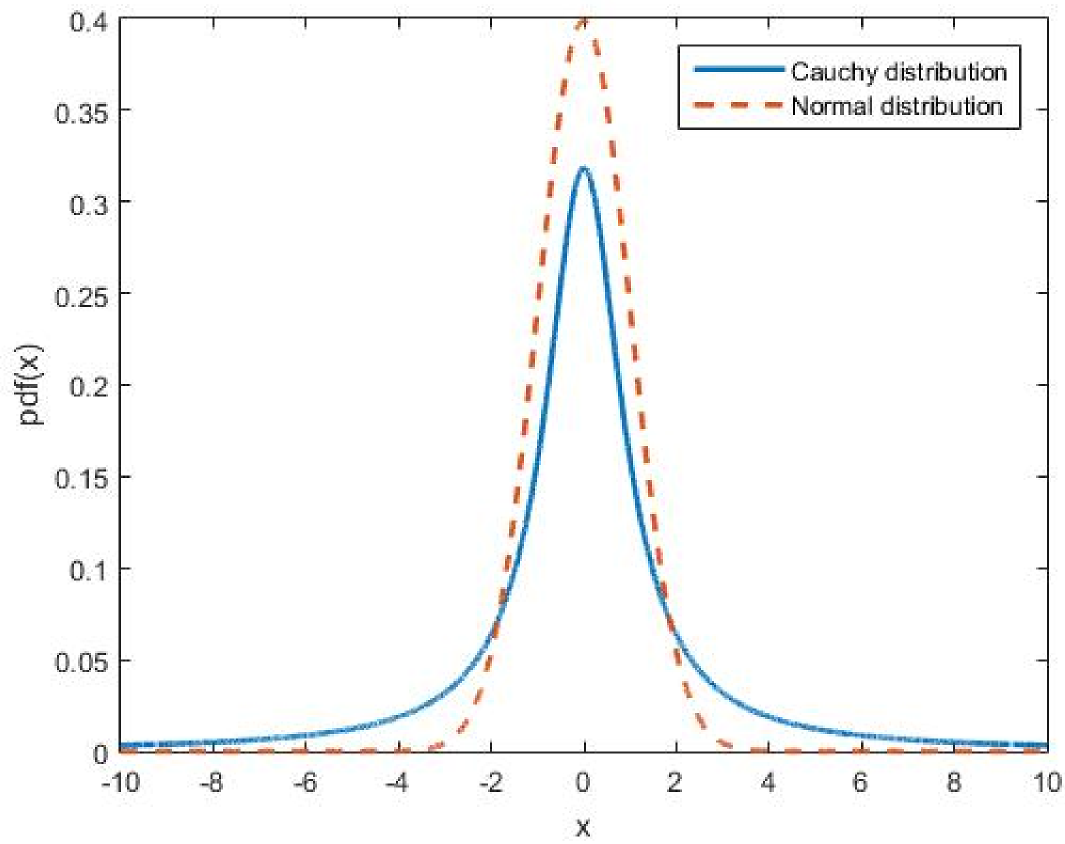

Figure 2: The probability density functions of the Cauchy distrubtion $S_1(1,0,0)$ and Standard Normal distribution $S_2(1,0,0)$. Note how the tail is "fatter" for smaller α.

Lemma 2.21. *Suppose X is a random variable with $\mathbb{E}[|X|^\gamma] < \infty$. Then for integers $0 \leq j \leq \gamma$, ϕ_X has finite derivative of order j given by $\phi_X^{(j)}(\theta) = \mathbb{E}[(iX)^j e^{i\theta X}]$.*

Proof. $j = 0$ is trivial. Assume that the statements holds for $j-1$, we wish to show that it holds for j. The result follows by induction if we differentiate under the integral sign for $j-1$. This is allowed since $\mathbb{E}[|X|^\gamma] < \infty$ and is proven in [24] proposition 9.2.1. □

We are are now ready to prove that α-stable Lévy processes exhibit fat tails for almost all α.

Proposition 2.22. *Suppose that the variance of an α-stable Lévy process is finite. Then $\alpha = 2$.*

Proof. Suppose $\alpha \in (0, 2)$ and that the variance, $\mathbb{E}[X^2] = \mathbb{E}[|X|^2]$ is finite. Then by lemma 2.21 the derivative of X must exist, however, we see by definition 2.19 that this is possible if and only if $\alpha = 2$. $\square$

In fact the result can be extended to the general tail behavior of Lévy stable processes and is done so in [23].

Proposition 2.23 (property 1.2.16 in [23]). *If X is a random variable with α-stable distribution for $\alpha \in (0, 2)$, then for any $\gamma \in (0, \alpha)$ $\mathbb{E}[|X|^\gamma] < \infty$, but $\mathbb{E}[|X|^\alpha] = \infty$.*

Alternate proof idea. A different way to think of it is to use fractional derivatives defined via $\phi_X^{(\gamma)} = \mathbb{E}[(iX)^\gamma e^{i\theta X}]$. Then the ideas of lemma 2.21 and proposition 2.22 might be extended to include the powers $\gamma \in (0, \alpha)$. Such an approach has not been spotted in the literature and may be interesting to investigate further.

$\square$

Thus not only do α-stable random variables have infinite variance for $\alpha < 2$ but their decay rate also decreases as the index of stability, α, increases. For $\alpha = 2$ we retrieve the Gaussian normal case which even has finite variance.

Example 2.8 (Cauchy distribution). *The Cauchy distribution with probability density function*

$$f(x) = \frac{1}{\pi(1 + x^2)} \tag{83}$$

is stable with $\alpha = 1$. We can find its characteristic function as

$$\begin{aligned} \phi_X(t) = \mathbb{E}[e^{itX}] &= \int_{-\infty}^{\infty} e^{itx} f dx = \int_{-\infty}^{\infty} e^{itx} \left(\frac{1}{\pi(1 + x^2)} \right) dx \\ &= e^{-|t|} \end{aligned} \tag{84}$$

by the Fourier inversion formula since the Fourier transform of $e^{-|t|}$ is precisely $\frac{2}{1+\omega^2}$. Compare this with definition 2.19 and the result follows.

We now restrict our attention to processes which are also symmetric. They are defined below.

Definition 2.20. *A random variable X is said to be stable symmetric (around 0) if its characteristic function is given by*

$$\phi_X(\theta) = \mathbb{E}[e^{i\theta X}] = e^{-\sigma^\alpha |\theta|^\alpha} \tag{85}$$

with α, σ as in definition 2.19.

Remark 2.20.1. *The attentive reader will note that the Cauchy distribution in the previous example fits this definition. So would also the normal distribution had we specified mean 0. This explains the parenthesis (around 0) in the definition above.*

We are now ready to once again set our class of processes in the context of fractal geometry.

Proposition 2.24. *An α-stable symmetric Lévy process is self-affine with self-affinity index $1/\alpha$.*

Proof. The increments $X(h) = X(t+h) - X(t)$ satisfy $\mathbb{E}[e^{i\theta(X(t+h)-X(t))}] = e^{-h\sigma^\alpha|\theta|^\alpha}$. Consider the characteristic functions of $X(ct)$ and $c^{1/\alpha}X(t)$ and note that they are the same. $\square$

Just as fractional Brownian motion illustrates the behavior of different-dimensional persistent behavior we use stable Lévy processes to better understand fractal dimension in the context of fat tails. The next proposition illustrates that the decay-rate of such fractals, and thus how fat their tails are, depends directly on the fractal dimension.

Finally, we are also able to show that the decay-rate as in proposition 2.23 is directly related to the fractal dimension of the graph.

Proposition 2.25. *The Hausdorff and Box dimensions of a symmetric α-stable Lévy process is a.s.* $\max(1, 2-1/\alpha)$.

Remark 2.25.1. *We will show below that* $\overline{D_B}$ graph $X \leq \max(1, 2-1/\alpha)$. *As the establishment of the lower bound is rather more difficult, we only outline a heuristic sketch of the idea. The derivations below are a slight extension of the partial proof found in [16].*

Proof of upper bound. We will show that $\overline{D_B}$ graph $X \leq \max(1, 2-1/\alpha)$ almost surely. Let R_f as in proposition 2.16. Since the process is self-affine with exponent $1/\alpha$ we have that

$$\mathbb{E}[R_X[t, t+\delta]] = \delta^{1/\alpha}\mathbb{E}[R_X[0,1]]. \tag{86}$$

By proposition 2.16 it then follows that the N_δ intersecting squares of the δ-mesh are bounded by

$$\mathbb{E}[N_\delta] \leq 2m + m\delta^{-1}\delta^{1/\alpha}\mathbb{E}[R_X[0,1]]. \tag{87}$$

where m is the least integer greater than or equal to $1/\delta$, so $m \leq 2/\delta$. It may be shown that $\mathbb{E}[R_X[0,1]] < \infty$. Hence $\mathbb{E}[N_\delta \delta^\beta] \leq c$ for suitable c and small δ with $\beta = \max(1, 2-1/\alpha)$. So

$$\mathbb{E}\left[\sum_{k=1}^{\infty} N_{2^{-k}}(2^{-k})^{\beta+\epsilon}\right] \leq c\sum_{k=1}^{\infty}(2^{-k})^{\epsilon} < \infty. \tag{88}$$

Therefore $N_\delta\delta^{\beta-\epsilon} \to < \infty$ almost surely. In particular, $N_\delta\delta^{\beta-\epsilon} \to 0$. The result follows by definition of the Hausdorff and Box-Counting dimensions. For more details consult propositions 4.1 and 16.8 of [16]. $\square$

Sketch of proof of lower bound. Let $\mu_f(A) = L\{t \in [0,1] : (f, f(t)) \in A\}$. This is the mass distribution needed for us to use proposition A.2. Then [16] defines the Fourier transform of μ_f as

$$\hat{\mu}_f(u) = \int_{\text{graph } f} e^{ix\cdot u} d\mu_f(x). \tag{89}$$

Notice that the s-potential as in definition A.2 is just the convolution

$$\Phi_s(x) = \int |x-y|^{-s} d\mu_f(y) = (|x|^{-s} * \mu_f)(x). \tag{90}$$

Then its Fourier transform is given by the product of the transforms of the elements of the convolution above. The transform of $|x|^{-s}$ is $c|u|^{s-2}$ where $c \in \mathbb{R}$ depends on s. Hence

$$\hat{\Phi}_s(u) = \hat{\mu}_f(u) c |u|^{s-2}. \tag{91}$$

Then by Parselval's theorem

$$\int \Phi_s(x) d\mu(x) = (2\pi)^2 \int \hat{\Phi}_s(u) \overline{\hat{\mu}_f(u)} du \tag{92}$$

and therefore

$$I_s(\mu_f) = (2\pi)^2 c \int |u|^{s-2} |\hat{\mu}_f(u)|^2 du. \tag{93}$$

The idea is now to use that (89) is closely related with the characteristic function as in definition 2.19 to establish that the integral (93) is finite if $\max(1, s \leq 2 - 1/\alpha)$ by proposition A.2.

□

Thus, finally we have shown that the higher the fractal dimension of certain random processes means that it is more wild. In the context of Lévy stable processes the decay-rate of the tails are characterized by the fractal dimension through proposition 2.23. Moreover, in the previous section it was shown that persistent and anti-persistent behavior of fractional Brownian motion is characterized by the fractal dimension. In fact, the degree of anti-persistent behavior was shown to be increasing with the fractal dimension. To us, it thus seems reasonable, in light of the Markowitz model, instead of variance, we try to minimize the fractal dimension. We now develop the tools necessary.

3 On the Empirical Estimation of the Fractal Dimension

So far all is well in theory, however, we need a practical tool to measure, or at least estimate, the dimension of a set. In this context a natural selection is the Hurst, or Hölder, exponent. The approach below stems from [25] and is an extension thereof. Consider a return series defined below:

Definition 3.1. *We define the time series of an asset, x_i, as a function $P^i(t) = P^i_t : \mathbb{I} \to \mathbb{R}$ that for any time value $t \in \mathbb{I}$ outputs the return, $P_i(t)$, of that asset, where $\mathbb{I}$ is any ordered set of return measurement points. Furthermore, when speaking of the fractal dimension such a time series, we implicitly refer to the fractal dimesion of the graph of $P^i(t)$.*

The Generalized Hurst Exponent of a (financial) time series, $\mathbb{H}_q$, is defined by

Definition 3.2. *The Generalized Hurst exponent is defined implicitly by*

$$\mathbb{H}_q \in \mathbb{R}^+ \ s.t. \ \left(\sum_{t=1}^{T} |P(t+\tau) - P(t)|^q\right)^{1/q} \propto \tau^{\mathbb{H}_q} \tag{94}$$

where $P(t)$ denotes the price at time t and $\tau \in \mathbb{R}^+$ denotes the step-distance.

The definition above is general, however, we will mainly focus on the case of where we have equality in (94) with $K = 1$ below:

$$\mathbb{H}_q = \frac{\ln\left(\sum_{t=1}^{T} |P(t+\tau) - P(t)|^q\right)^{1/q} - \ln K}{\ln \tau}. \tag{95}$$

Remark 3.2.1. *This form (with $K = 1$) is namely invariant under affine transformations of the time series since any constants (multiplicative or additive) are cancelled. This affine invariance mirrors the bi-Lipschitz invariance of the Hausdorff dimension shown in theorem 2.15.*

In the following sections we will build on the established theory for the exponent. The convergence properties of the exponent, in one form or another, date back to the last century. We will build on these properties, found in the section immediately below, to establish the general result which ties together the Box-Counting dimension of a convex combination of functions.

3.1 Limit Approximations for the Fractal Dimension

Proposition 3.1. *Let $X, Y \subset \mathbb{R}$. Then for a graph of a piecewise continuous function $f : X \to Y$ it holds that $\lim_{\tau\to 0^+} \mathbb{H}_1 = 2 - D_B$. I.e. the first moment Hurst exponent is asymptotically accurate in measuring the Box-Counting dimension.*

Proof. We can without loss of generality assume that the total time measurement length is 1. Let $R_\tau = \langle |P_{t+1} - P_t| \rangle$ where $\tau = \frac{1}{2^m}, m \in \mathbb{N}$ is the step length, such that there are $\frac{1}{\tau} = 2^m$ subintervals. If $\mathbb{H}_1^*$ denotes the asymptotic first moment Hurst exponent we have that

$$\mathbb{H}_1^* = \lim_{\tau\to 0} \frac{\ln R_\tau - \ln K}{\ln \tau}. \tag{96}$$

Where K is a constant of proportionality attributed to (94). Clearly $\lim_{\tau\to 0} \frac{\ln K}{\ln \tau} = 0, \forall K \in \mathbb{R}$. We can thus, without loss of generality, disregard $\ln K$ in the following limit argument.

Next, by piecewise continuity of f there will be a mean of $\frac{R_\tau}{\tau}$ boxes in each subinterval of length τ, and thus, since there are $\frac{1}{\tau}$ subintervals, there are $\frac{R_\tau}{\tau^2}$ boxes, implying $\tau^2 N_\tau = R_\tau$.

Therefore

$$\mathbb{H}_1^* = \lim_{\tau\to 0} \frac{\ln R_\tau}{\ln \tau} = \lim_{\tau\to 0} \frac{\ln \tau^2 N_\tau}{\ln \tau} = \lim_{\tau\to 0} \frac{\ln N_\tau + \ln \tau^2}{\ln \tau} = 2 - D_B \tag{97}$$

as desired.

□

Before generalizing the above proposition to higher moments of q, we need to do some extra leg-work. It is useful to define $\tau_0 = \min(1, \sup\{x \in \mathbb{R} : |P(t+x) - P(t)| < 1, x > 0\})$. This serves several purposes. First and foremost it guarantees that for $\tau \in (0, \tau_0)$ we have $\ln\langle|P_{t+1} - P - t|^q\rangle < 0$ for all $q \geq 1$. Further it guarantees that $\langle|P_{t+1} - P - t|\rangle \geq \langle|P_{t+1} - P - t|^q\rangle$ for all $q \geq 1$. This will make the following lemma next to tautological.

Lemma 3.2. *For $\tau \in (0, \tau_0)$, $\mathbb{H}(q) = \mathbb{H}_q$ for any time series is an increasing function of q.*

Proof. From (95) we immediately see that for fixed $\tau < \tau_0$, $\mathbb{H}_q$ increases as q increases since K only depends on τ. □

We are now prepared to generalize the previous proposition to higher moments of the Hurst exponent. This step's usefulness relates to the more beneficial smoothness properties higher orders possess.

Proposition 3.3. *Let $X, Y \subset \mathbb{R}$. Then for a graph of a piecewise continuous function $f : X \to Y$, it holds that $\lim_{\tau\to 0^+} \mathbb{H}_q = 2 - D_B$. I.e. every moment of the Hurst exponent is asymptotically accurate in measuring the Box-Counting dimension.*

Proof. We can without loss of generality assume that the total measurement length is 1. Let $R_\tau = \langle|P_{t+1} - P_t|^q\rangle^{1/q}$, such that there are $\frac{1}{\tau}$ subintervals.

Instead of using the box definition of Box-Counting definition, we use the equivalent definition where instead we use the largest set of rectangles of size $\inf |P_{t+1} - P_t|^q$ by τ. In case there does not exists $\varepsilon > 0$ such that for all $\tau \in N_\varepsilon(0) \cap \mathbb{R}^+$, $\tau > \inf |P_{t+1} - P_t|^q$ we can always rescale the time series (which does not change the dimension by dimensional invariance under affine transformations as shown in theorem 2.15) such that the piece is sufficiently small (in height).

Thus for a piecewise continuous function, for which we have already established existence of D_B, we have that

$$\limsup_{\tau\to 0^+} -\ln \tau^2 N_\tau \geq -\ln\left(\liminf_{\tau\to 0^+} R_\tau\right). \tag{98}$$

Whereupon it is established that

$$2 - D_B = 2 - \limsup_{\tau\to 0^+} \frac{\ln N_\tau}{-\ln \tau} = \limsup_{\tau\to 0^+} \frac{\ln N_\tau \tau^2}{\ln \tau} \geq \limsup_{\tau\to 0^+} \frac{R_\tau}{\tau} = \mathbb{H}_q^*. \tag{99}$$

However, by the lemma above,

$$\mathbb{H}_q \geq \mathbb{H}_w \tag{100}$$

for $q \geq w \geq 1$, $\tau \in (0, \tau_0)$ and by the previously established fact that

$$\mathbb{H}_1^* = 2 - D_B \tag{101}$$

we conclude that

$$\mathbb{H}_q^* = 2 - D_B. \tag{102}$$

□

Remark 3.3.1. *It would be interesting to investigate whether the piecewise continuity assumption above can be replaced by $f \in L^p(X)$, for instance for $p = q$. In practice, a time series will always be represented by a continuous function since the measurement points are finite. However, it may be of theoretical interest as this might allow us to take convex combinations of sets like the Cantor dust and study their dimension. That is, we would render it onto the plane as the graph of a function.*

3.2 Concavity of the Box-Counting Dimension

The Hurst exponent is a measure of jaggedness and long-term memory of the time series. It is thus closely related with the Box-Counting dimension of certain processes as discussed in section 2. As we have motivated in that section, a high fractal dimension is undesirable for an asset that we wish to invest in since it may exhibit fat tails or autocorrelation. It is therefore natural to try to minimize such an undesirable property. In order to do so, we prove that the Hausdorff dimension when taken over a convex combination of functions, is a concave function. This approach is to the best of our knowledge novel.

Proposition 3.4. *For all $1 > \tau > 0$, $\mathbb{H}_1(x)$ of a financial time series is a convex function of the portfolio vector, x.*

Proof. We replace proportionality with an arbitrary constant, K, and set $q = 1$ and sum notation is introduced. Thus

$$K\tau^{\mathbb{H}_1} = \sum_{t \in I} |P_{t+1} - P_t|. \tag{103}$$

As we are interested in portfolio returns and not those of individual time series, $P(\cdot) = \sum_{i=1}^n x_i P^i(\cdot)$. So

$$\begin{aligned} K\tau^{\mathbb{H}_1(x)} &= \sum_{t \in I} |\sum_{i=1}^n x_i(P_{t+1}^i - P_t^i)| \\ \Rightarrow \mathbb{H}_1(x) &= \frac{\ln\left(\sum_{t\in I} |\sum_{i=1}^n x_i(P_{t+1}^i - P_t^i)|\right) - \ln K}{\ln \tau}. \end{aligned} \tag{104}$$

The result now follows from the composition rules of convex and concave functions since $\mathbb{H}_1(x)$ is the negative (since τ is the inverse of the number of steps and thus $\tau < 1$) logarithm (convex) of the sum (affine) of the absolute values (convex) of the sum of affine functions and each is increasing.

□

Sadly, $\mathbb{H}_1$ is only piecewise differentiable without additional assumptions. We therefore turn our attention to the generalized Hurst exponent for even $q \geq 2$.

Proposition 3.5. *For all $1 > \tau > 0$, and even $q \geq 2$, $\mathbb{H}_q(x)$ of a financial time series is a convex, q-times continuously differentiable function of the portfolio vector, x.*

Proof. We replace proportionality with an arbritrary constant, K and sum notation is introduced. Thus

$$K\tau^{\mathbb{H}_q} = \left(\sum_{t\in I} |P_{t+1} - P_t|^q\right)^{1/q}. \tag{105}$$

As we are interested in portfolio returns and not those of individual time series, $P(\cdot) = \sum_{i=1}^n x_i P^i(\cdot)$. Further, since q is even we can remove the absolute value function, such that

$$\begin{aligned} K\tau^{\mathbb{H}_q(x)} &= \left(\sum_{t\in I}(\sum_{i=1}^n x_i(P_{t+1}^i - P_t^i))^q\right)^{1/q} \\ \Rightarrow \mathbb{H}_q(x) &= \frac{\ln\left(\sum_{t\in I}(\sum_{i=1}^n x_i(P_{t+1}^i - P_t^i))^q\right)^{1/q} - \ln K}{\ln\tau}. \end{aligned} \tag{106}$$

The result now follows from the composition rules of convex and concave functions since $\mathbb{H}_q(x)$ is the negative logarithm (convex) of the qth root (convex for even q) of the sum (affine) of qth powers (convex) and each is increasing. Note further that all these compositions are at least q times continuously differentiable, thus $\mathbb{H}_q(x)$ is too.

□

Even though at first glance we would expect an averaging process to make the data more smooth, these results should not surprise our intuition. They simply state that the weighted average of several time series is more jagged than its most jagged component series. Heuristically, this happens since we should not expect the different fractal shapes to "cancel" since they are in a sense the product of an infinitely limiting process. We now turn to differentiability, our interest thereof should be apparent in regard to optimization. Moreover, if the reader is not already convinced by the strength of these results, we also have the following more general corollary which rests on the convexity properties of the Hurst exponent.

Proposition 3.6. *Let $X, Y \subset \mathbb{R}$. The Box-Counting dimension of a weighted average of any number, $n \in \mathbb{N}$, of piecewise continuous functions $f_i : X \to Y$ is a concave function of the weightings x_i. I.e.*

$$\sum_{i=1}^n x_i D_B(f_i) \leq D_B\left(\sum_{i=1}^n x_i f_i\right). \tag{107}$$

Where $\sum_{i=1}^n x_i = 1,\ x_i \geq 0,\ i = 1, 2, ..., n.$

Proof. This is an immediate corollary of the convexity of $\mathbb{H}_1$, the preservation of convexity under limits, and the asymptotic equality between $\mathbb{H}_1$ and D_B.

□

We end this section with an example to illustrate the meaning of these ideas.

Example 3.1. *Consider an $n+m$-tuple of n fBm processes, B_i, of different dimension $D_{B,i}$ and m symmetric stable Lévy processes, L_j, of dimensions $D_{B,j}$. Then the dimension of the graph of*

$$\sum_i^n x_i B_i + \sum_j^m y_j L_j \tag{108}$$

is a concave function of (x,y) where $x = (x_1, ..., x_n) \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ is defined equivalently.

The example is of course also true for the Hausdorff dimension as it is equal to the Box-Counting dimension for the processes considered.

4 Optimal Portfolios for a Fractal Measure of Risk

Before we begin with the solution of the problem, we note a particularly useful form of $\mathbb{H}_2$.

Proposition 4.1. *Let*

$$\mathcal{M} = \sum_{t\in\mathbb{I}} \Delta P(t)\Delta P^t(t). \tag{109}$$

Then

$$\begin{aligned} \mathbb{H}_2(x) &= \frac{\ln\left(x^t\mathcal{M}x\right) - \ln K}{2\ln\tau} \quad \textit{and} \\ \nabla\mathbb{H}_2(x) &= \frac{\mathcal{M}x}{(\ln\tau)x^t\mathcal{M}x}. \end{aligned} \tag{110}$$

Proof. We have, assuming $K = 1$ without loss of generality, since we will differentiate away the constant anyway. Now let $\Delta P(t) = (P_1(t+1) - P_1(t), ..., (P_n(t+1) - P_n(t)))^t$ then

$$\begin{aligned} \mathbb{H}_q(x) &= \frac{\ln\left(\sum_{t\in I}(\sum_{i=1}^n x_i(P^i_{t+1} - P^i_t))^q\right)^{1/q}}{\ln\tau} = \frac{\ln\left(\sum_{t\in I}(x^t\Delta P(t))^q\right)}{q\ln\tau}, \\ \mathbb{H}_2(x) &= \frac{\ln\left(\sum_{t\in I}(x^t\Delta P(t))^2\right)}{2\ln\tau} = \frac{\ln\left(\sum_{t\in I}(x^t\Delta P(t))(x^t\Delta P(t))\right)}{2\ln\tau} \\ &= \frac{\ln\left(\sum_{t\in I}(x^t\Delta P(t))(\Delta P^t(t)x)\right)}{2\ln\tau} = \frac{\ln\left(\sum_{t\in I} x^t(\Delta P(t)(\Delta P^t(t))x\right)}{2\ln\tau} \\ &= \frac{\ln\left(x^t\sum_{t\in I}\left[(\Delta P(t)(\Delta P^t(t))\right]x\right)}{2\ln\tau} = \frac{\ln\left(x^t\mathcal{M}x\right)}{2\ln\tau}. \end{aligned} \tag{111}$$

And by the chain rule and differentiation of quadratic forms the result follows. □

We are now ready to present the main problem of this project. Namely to minimize (a heuristic measure of) the fractal dimension for a portfolio.

Problem 5.

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} -\frac{\ln\left(x^t\mathcal{M}x\right)}{2\ln\tau} - \mu\bar{r}^t x, \\ &s.t.\ e^t x = 1, \\ &\qquad x \geq 0. \end{aligned} \tag{112}$$

The constraint $x \geq 0$ precisely means that short-selling is banned. Further, since this makes the feasible set compact, we are guaranteed the existence of a minimum. In fact, it is trivially found under suitable assumptions on the objective function since the feasible set is a polyhedral set.

4.1 Portfolios of Minimal Dimension

Proposition 4.2. *$\mathcal{M}$ is positive semi-definite.*

Proof. Let $x \in \mathbb{R}^n$ and consider $x^t\mathcal{M}x$. We can write this as

$$x^t\mathcal{M}x = x^t\sum_{t\in\mathbb{I}}\Delta P(t)\Delta P^t(t)x = \sum_{t\in\mathbb{I}} x^t\Delta P(t)\Delta P^t(t)x = \sum_{t\in\mathbb{I}}(x^t\Delta P(t))^2 \geq 0. \tag{113}$$

□

Assumption 3. *$\mathcal{M}$ is positive definite.*

We do not need the assumption above for the solution of problem 5 but it is necessary for completeness of the theory. If we compare with the problems found in the section on Markowitz's original model, the PD assumption was necessary to find a solution since we used matrix inversions. Even if we do not work with matrix inversions here, this may be useful in completing the theory with alternate objective function approaches that still include the Hurst exponent.

Proposition 4.3. *The objective function (112) is concave and strictly-concave under assumption 3.*

Proof. Using the gradient obtained in (111) we find that the assoicated Hessian matrix is

$$-\frac{\mathcal{M}\ln\tau x^t\mathcal{M}x - 2\mathcal{M}x^t\mathcal{M}x\ln\tau}{(\ln\tau x^t\mathcal{M}x)^2} = \frac{\mathcal{M}}{(\ln\tau x^t\mathcal{M}x)} \tag{114}$$

which is a negative multiple of $\mathcal{M}$ (remember $\ln\tau < 0$). The result follows by proposition 4.2. □

The concavity shown above allows us to transfer the applied problem of finding an optimal portfolio to that of finding the extreme point of a particularly simple convex polyhedron.

Proposition 4.4. *Problem 5 is solved by some $x \in C_E$ where*

$$\begin{aligned} C_E =& \bigcup_{i=1}^{n} \{x \in C : x_i = 1, x_j = 0 \, \forall j \neq i\}, \\ C =& \{x \in \mathbb{R}^n : x \geq 0, e^t x = 1\}. \end{aligned} \tag{115}$$

Proof. This is an immediate corollary of theorem 2.8 and proposition 2.9. □

Remark 4.4.1. *A less complicated way to say this would be to say that problem 5 is solved by one of the standard Euclidean basis vectors e_i.*

We have thus reduced problem 5 to evaluating n functions and choosing the smallest. This can be done numerically in MATLAB in next to no time. Moreover, this illustrates an interesting property of the concavity of the dimension of portfolio vis-à-vis the convexity of the variance. Loosely speaking, the superposition of time-series of different dimension actually increases their dimension, whereas it would decrease the variance. The implication is that investors following the simple decision rule specified by the optimization problem 5 will simply the pick the one asset which suits their relative preference of jaggedness versus return whereas the Markowitz investor would prefer to "spread his risks". We do not think it realistic to contradict the old adage of not putting all eggs in one basket, but it is nevertheless interesting to investigate how well the specification does in actually picking individual stocks. Another route, to avoid this, might be to "convexify" the objective function by composing it with the exponential function or to add a variance term to it. Even better fated might be the attempt to convexify and then add a variance term (or in the opposite order). We discuss this further in the next section.

4.2 Alternate Objective Functions

Regrettably the concavity of the objective function of problem 5 forces the optimal solution to a basis vector e_i. This means that the "optimal portfolio" in our sense is completely undiversified. We therefore present 2 alternative objective functions f_1, f_2 and comment briefly on them.

The first which comes to mind is the simple "convexification" discussed above

$$f_1 = \exp\left(k\mathbb{H}_2(x)\right) - \mu \bar{r}^t x. \tag{116}$$

(116) has the advantage that for suitable choice of k it is convex. Moreover, in the extreme case of $\mu = 0$ it even has the same minimum as problem 5 since $\exp(\cdot)$ is increasing on the feasible set. However, there is no real justification for selecting this objective function other than that it has nicer properties then the original one. Problem 5 is clearly not equivalent with the alternation using f_1.

The alternative approach is to simply add a variance term to objective function of problem 5. We obtain

$$f_2 = \nu \frac{x^t \Sigma x}{2} - \mathbb{H}_2(x) - \mu \bar{r}^t x. \tag{117}$$

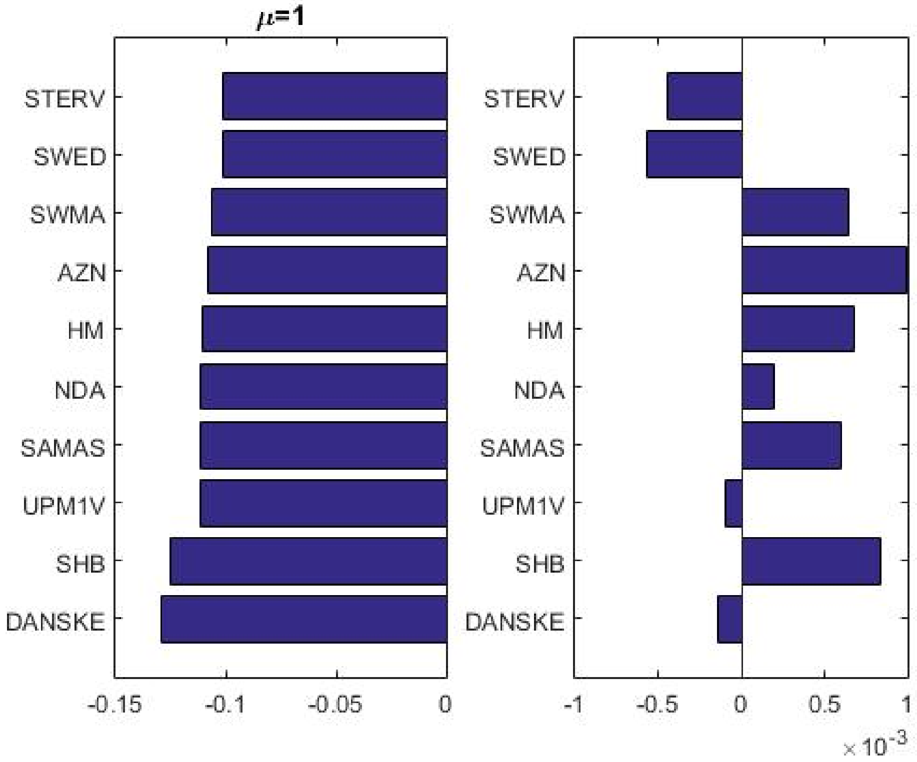

Figure 3: The left graph shows the 10 smallest objective functions for vectors e_i (these correspond to the investment strategies of investing purely in one stock). The right graph shows the difference between the return of the 10 "most optimal" (in regard to the optimization problem 5) stocks and the Markowitz portfolio as given by the index. Estimation was carried out between the first of January 2004 and the first of January 2008. Returns were measured between the first of January 2008 and the first of January 2010. Another figure can be found on the next page.

f_2 is not convex nor concave. However, it falls under the broader class of difference of convex (DC) functions. There exist convergent algorithms for finding optimum of such functions and they are considered in [26] in more detail.

4.3 Empirical tests

We now present data from the OMX Nordic 40 index (data retrieved from [27] and [28]) with which we compute numerically the optimal portfolio of problem 5. This is almost trivial due to proposition 4.4. Our intention is to compare our portfolio selection model with that of Markowitz. However, according to the Capital Asset Pricing Model (CAPM) introduced by Sharpe in [29] we need not actually calculate Markowitz's portfolio (problem 3 with an added non-negativity constraint $x \geq 0$) since it is equivalent to

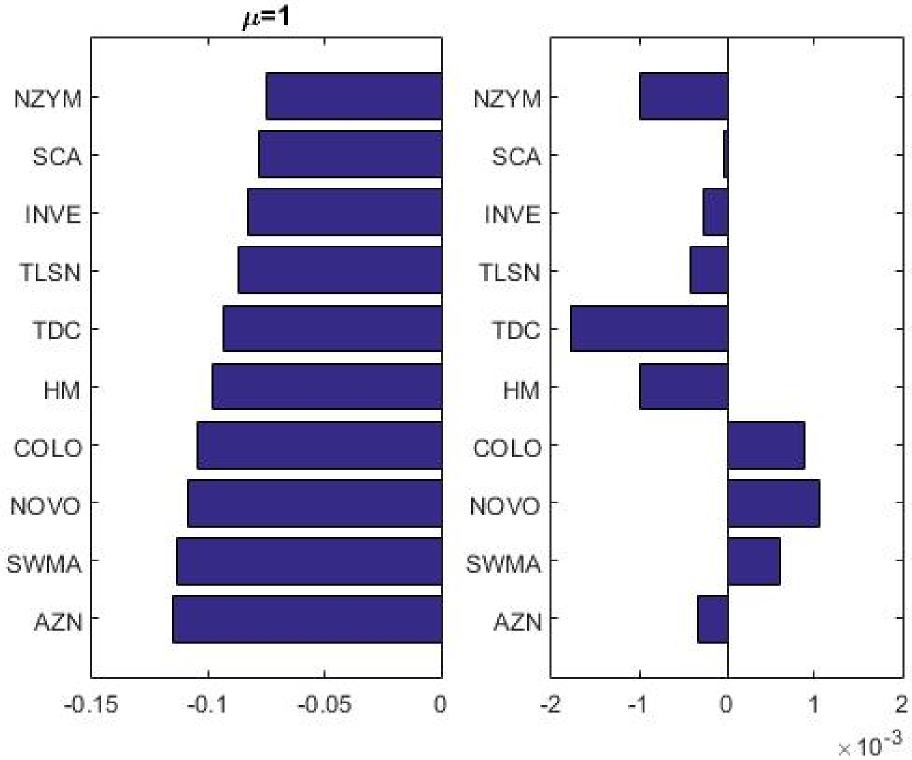

Figure 4: The same as the preceding two graphs but for different times. Estimation was carried out between the first of January 2008 and the first of January 2010. Returns were measured between the first of January 2010 and the first of January 2012.

the market portfolio. We remind the reader that the market portfolio is the weighted average of all stocks combined in one portfolio where the weighting is according to their relative market capitalization (the value of the company considered). Since we only consider the Nordic market, our market portfolio will precisely be the OMX Nordic 40 index and we therefore compare the returns of that index with the stocks optimal according to the solution of problem 5.

The procedure for testing our model is done as follows. We compute the objective function for all 40 stocks in the index, which we then compare for $\mu = 1$ where data is taken from an estimation period (for instance between 2004 and 2008). We do not justify the assumption $\mu = 1$. This would be part of calibrating the model somehow for preferences of risk and return. Nevertheless, it can be noted that the ordering does not change much for different values of μ and that this can easily be checked with the code included. Finally we compare the returns of our 10 best stock picks with that of the index for a trial period (e.g. 2008-2010).

In reference to the figures 3 and 4, it is observed that our 10 stocks, at least on average, outperformed the market for the first trial period as found in figure 3. The converse is true for the second trial period.

An explanation for this may be that the first trial period overlaps with the recent financial crisis. The importance of non-turbulent stock picks may thus be greater than otherwise, i.e. if the market is stable. Thus, the relative safety of low-dimensional stocks could be beneficial during wild market movements whereas they just exhibit a further cost when the market is calm. It must also be recognized that the amount of data presented is insufficient to draw any definite conclusions. Nevertheless, we believe its inclusion to be illustrative of the fact that our measure is easy to work with numerically.

5 Discussion and Conclusion

In this thesis we have presented results from optimization theory, portfolio theory, probability and fractal geometry in order investigate portfolio theory under heavy tail and autocorrelation assumptions. We used saddle-point optimality to derive the Kurosh-Kuhn-Tucker conditions for optimality which we then applied to solve Markowitz's portfolio model. In order to expand on his work, we illustrated the importance of fractal dimension (here the Hausdorff, or Box-Counting, dimensions) with respect to fractional Brownian motion and symmetric α-stable Lévy processes. Specifically, we showed for these processes that a higher fractal dimension implies undesirable properties for investment. Using this, we then motivated and constructed a suitable objective function for minimization. We solved this problem in section 4 by characterizing the candidates for optimality by the extreme points of the convex polyhedron $C = \{x \in \mathbb{R}^n : x \geq 0, e^t x = 1\}$ which precisely are the Euclidean basis vectors e_i. Further, we gave an example through the data acquired from the Nasdaq OMX Nordic 40 index and compared our results to those of Markowitz.

Our main contributions are the concavity of the Box-Counting dimension, D_B, of propositions 3.4 to 3.6 but also the statement and solution of problem 5. Lastly, we also have some thoughts on how to improve the framework in the following.

Hurst, or as in this context more suitable, Hölder, exponents started out as a local property of a set, due to Hölder. Our thinking is in line with that of fractional Brownian motion (or symmetric Lévy stable motion), which has one Hölder exponent, as demonstrated, globally. However, real financial markets generally have a multitude of local Hölder exponents. We will often note in time series that where variation is the wildest, Lebesque measure is the smallest. This leaves us vulnerable to averaging out these wild fluctuations which will, to some extent, make us miss the point of finding disaster risk which should be the greatest precisely on these sets of small Lebesque measure.

Problem 5 may perhaps be generalized to the broader study of multifractals. Mandelbrot et al. illustrate in [19] that multifractals have a time-dependent Hölder exponent, $\tau(t)$, which is concave (as a function of time). We could thus perhaps consider maximizing the integral

$$\int (\tau(t, x) + \rho(t, x))\, dtdx \tag{118}$$

or some variation thereof. This author speculates that the theories of optimal control and the calculus of variations may well be useful in this study. Moreover, the objective functions f_1, f_2 as described in (116,117) may also be of interest for further research. In particular, some variation of (117) may be interesting as it is a mixture of our and existing theory. Beside the fact that it accounts for the diversification issue discussed previously, it might help shed further light on the issue whether our risk measure works well during turbulent times.

A Further Computational Techniques for Dimensions

These topics are sometimes far from the main issue of this thesis and we believe them to disturb the flow of ideas. In order to precisely determine the fractal dimension of fractional Brownian motion we need some basic potential theory. The approach stems from chapter 4.3 of [16].

Definition A.1. *Let μ be a measure defined on a bounded subset X of $\mathbb{R}^n$. If $0 < \mu(X) < \infty$, we say that μ is a mass distribution.*

Example A.1. *Any probability measure on a bounded set $X \subset \mathbb{R}^n$ is a mass distribution since $\mu(X) = 1$.*

The example above sets us in the right direction (considering fBm). We first construct a lower bound for the Hausdorff measure in order to proceed to construct a lower bound on the dimension.

Proposition A.1 (Proposition 4.9a of [16])**.** *Let μ be a mass distribution, let $F \subset \mathbb{R}^n$ be a Borel set, and let $0 < c < \infty$. Then if $\limsup_{r\to 0} \mu(N_r(x))/r^s < c$ for all $x \in F$ then $\mathcal{H}^s(F) \geq \mu(F)/c$.*

Proof. For $\delta > 0$ let

$$F_\delta = \{x \in F : \mu(N_r(x)) < cr^s \text{ for all } 0 < r \leq \delta\}. \tag{119}$$

Next, let $\{U_i\}$ be a δ-cover of F. Then by definition of F_δ it is also a cover of that set. Then for a neighborhood $N_{\text{diam }(U_i)}(x)$ such that $U_i \cap F_\delta \neq \emptyset$ that N certainly contains U_i. Hence,

$$\mu(U_i) \leq \mu(N) \leq c(\text{diam }(U_i))^s. \tag{120}$$

Thus

$$\mu(F_\delta) \leq \sum_i \{\mu(U_i) : U_i \cap F_\delta \neq 0\} \leq c \sum_i (\text{diam }(U_i))^s. \tag{121}$$

Since $\mu(U_i)$ is any δ-coveer of F, it follows that $\mu(F_\delta) \leq c\mathcal{H}^s_\delta(F) \leq c\mathcal{H}^s(F)$. Then set $y = 1/\delta$. The sets F_y are then increasing so the limits can be moved inside the measure and the result follows. □

Definition A.2. *For $s \geq 0$ the s-potential at a point $x \in \mathbb{R}^n$ of the mass distribution μ defined on $X \subset \mathbb{R}^n$ is defined as*

$$\Phi_s(x) = \int_X \frac{d\mu(y)}{|x-y|^s}. \tag{122}$$

The associated s-energy is

$$I_s(\mu) = \int_X \Phi_s(x) d\mu(x). \tag{123}$$

We are now ready to use the integrals above to find a lower bound on the Hausdorff dimension. This is the result used to find the lower bound of the Hausdorff dimension of the graph of fBm. The idea is that if the energy is finite, the distribution of mass must be finite almost everywhere. We then use the previous proposition to extract the result desired.

Proposition A.2 (Theorem 4.13a of [16])**.** *Let X be a subset of $\mathbb{R}^n$. If there is a mass distribution μ on X with $I_s(\mu) < \infty$ then $\mathcal{H}^s(X) = \infty$ and $D_H(X) \geq s$.*

Proof. Let μ be a mass distribution on X and suppose that $I_s(\mu) < \infty$. Also let

$$E = \left\{ x \in X : \limsup_{r \to 0} \mu(N_r(x))/r^s > 0 \right\}. \tag{124}$$

Then for $x \in E$ there exists strictly decreasing $\{r_i\} \to 0$ such that $\mu(N_{r_i}(x)) \geq \epsilon r_i^s$. Now define $q_i = (r_i + r_{i+1})/2$ such that $0 < r_{i+1} < q_i < r_i$. If we let $A_i = N_{r_i}(x) \setminus N_{q_i}(x)$ we get $\mu(A_i) \geq \frac{1}{4}\epsilon r_i^s$ by continuity. Thus

$$\Phi_s(x) = \int_X \frac{d\mu(y)}{|x-y|^s} \geq \sum_{i=1}^{\infty} \int_{A_i} \frac{d\mu(y)}{|x-y|^s} \geq \sum_{i=1}^{\infty} \frac{1}{4}\epsilon = \infty \tag{125}$$

since $|x-y|^{-s} \geq r_i^{-s}$ on A_i. However since the s-energy is finite the mass must be distributed on a set of measure 0 and $\Phi_s(x) < \infty$ almost everywhere. Hence $\mu(E) = 0$. Therefore, $\limsup_{r \to 0} \mu(N_r(x))/r^s = 0$ for $x \in X \setminus E$. The result follows by applying the previous proposition. □

B MATLAB Code

B.1 Index Data Management

```
% read_files_into_big_matrix.m
clear all
close('all')

names=dir('*.dat');  % directory listing

d=zeros(6000,length(names)+1);  % allocate big matrix

start=datenum('2000-01-01');

for i=1:length(names)
   % disp(names(i).name)
   clear q
   q=importdata(names(i).name,';');  % stuff into column of matrix
   datum=datenum(q.textdata)-start;
```

```
   value=q.data;
   for j=1:length(q.data)
       d(datum(j),i)=value(j);
       if (i==1)
           d(datum(j),length(names)+1)=datum(j)+start;
       end
   end
end

jj=0;
for j=1:6000
    if (d(j,41)>3000)
        jj=jj+1;
        dd(jj,:)=d(j,:);
    end
end

dataarray=datestr(dd(:,41));


% load_omx_data.m
d=zeros(6000,2);  % allocate big matrix
q=importdata('omx/OMX.dat',';');

start=datenum('2000-01-01');
datum=datenum(q.textdata)-start;
value=q.data;
for j=1:length(q.data)
    d(datum(j),1)=value(j);
    d(datum(j),2)=datum(j)+start;
end

omx=zeros(10,2);
jj=0;
for j=1:6000
    if (d(j,2)>3000)
        jj=jj+1;
        omx(jj,:)=d(j,:);
    end
end

omxdate=datestr(omx(:,2));

% all returns
```

```
omx_r=(omx(2:end,1)-omx(1:end-1,1))./omx(1:end-1,1);

% function for average return between T1 and T2
omx_R=@(T1,T2)mean(omx_r(T1:T2))';
```

B.2 Computations

```
allstock=zeros(2912,40);
allreturn=zeros(2913,40);
R=zeros(40,1);

for stock=1:40

tim=dd(1:end-1,41)-dd(1,41);
r=(dd(2:end,stock)-dd(1:end-1,stock))./dd(1:end-1,stock);

for k=1:length(r)
    if isnan(r(k))
        r(k)=0;
    end
    if r(k) < -0.99 | r(k)> 5000
        r(k)=0;
    end
end

if 0
    subplot(2,1,1); plot(tim,r)
    subplot(2,1,2); hist(r,50)
end

allreturn(:,stock)=r;

dP=r(2:end)-r(1:end-1);

% plot if necessary
if 0
    plot(dP)
    axis([0 3000 -1 1])
    title(num2str(stock))
    pause(0.1);
end

allstock(:,stock)=dP;    % big matrix
```

```
end

% covariance matrix for the entire time 1 to 2912
M=allstock'*allstock;

% function to get covariance matrix between T1 and T2
MMM=@(T1,T2)allstock(T1:T2,:)'*allstock(T1:T2,:);

% function to get average return
R=@(T1,T2)mean(allreturn(T1:T2,:))';

% Hurst exponent of portfolio vector x
H2=@(x,T1,T2)(-0.5*log(x'*MMM(T1,T2)*x)/log(T2-T1));
%we are not guaranteed non-neg since we assumed K=1

% return of portfolio vector x
RR=@(x,T1,T2)x'*R(T1,T2);

if 0
    hurst=zeros(40,1);
    ret=zeros(40,1);
    T1=500;
    T2=2900;
    for k=1:40
        x=zeros(40,1);
        x(k)=1;
        hurst(k)=H2(x,T1,T2);
        ret(k)=RR(x,T1,T2);
    end
    hurst_ret=[hurst, ret]
    subplot(2,1,1); bar(hurst);
    subplot(2,1,2); bar(ret);
end

cost=@(x,T1,T2,mu)(-H2(x,T1,T2)-mu*RR(x,T1,T2));

%.....get ticker name
for k=1:40
    company(k)=cellstr(names(k).name(1:end-4));
end

if 1
    % close('all')
```

```
    omxx=omx_R(2037+600,2548+600)             %..omx return

    %Change T1,T2 etc in the script for different graphs and also omxx
    %above.

    for mu=1;
    T1=1657;
    T2=2165;
%    mu=0.1;
    obj0=zeros(40,1);
    for k=1:40
        x=zeros(40,1);
        x(k)=1;
        obj0(k)=cost(x,T1,T2,mu);
    end
    % bar(obj0)
    [s,indx]=sort(obj0,'ascend');  % indx contains where they come from
    index=indx(2:11);
    if 1
        subplot(1,2,1);
        barh(obj0(index))
        title(['\mu=' num2str(mu)])
    end
    % index

    set(gca,'YTick',1:length(index),'Ylim',[0 length(index)+1])  % ticker axis labels
    set(gca,'YTickLabel',company(index))

    TT1=1657;   % future time
    TT2=2165;
    ret_later=R(TT1,TT2);
    hitlist=[index, ret_later(index)];
    if 1
        subplot(1,2,2);
    end
    barh(ret_later(index)-omxx)

    set(gca,'YTick',1:length(index),'Ylim',[0 length(index)+1])
    set(gca,'YTickLabel',company(index)) % ticker axis labels

    pause(0.001)
    end
```

```
    if 0    % scatter plot
    figure
    plot(obj0,ret_later,'*')
    title('Correlation Objective function vs. later return')
    xlabel('Objective function')
    ylabel('Later return')
    end
end
```

References

[1] J. Fourier. *Théorie Analytique de la Chaleur.* Chez Firmin Didot, Paris, 1822.

[2] B. Mandelbrot. *The Fractal Geometry of Nature.* Freeman, New York, 1982.

[3] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952.

[4] F. Knight. *Risk, Uncertainty and Profit.* Hart, Schaffner & Marx, New York, 1921.

[5] L. Bachelier. *Théorie de la Speculation.* PhD thesis, University of Paris, 1900.

[6] B. Mandelbrot. *The Misbehavior of Markets.* Basic Books, New York, 2004.

[7] E. Fama and K. French. Size, value and momentum in international stock returns. *Journal of Financial Economics*, 105(3):457–472, 2012.

[8] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.

[9] W. Rudin. *Foundations of Mathematical Analysis.* McGraw-Hill, New York, 1976.

[10] A. Friedman. *Foundations of Modern Analysis.* Dover, New York, 1982.

[11] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, Singapore, 1987.

[12] A. Pinkus and S. Zafrany. *Fourier Series and Integral Transforms.* Cambridge University Press, Cambridge, 1997.

[13] M. Bazaraa, Sherali H., and Shetty C. *Nonlinear Programming.* John Wiley & Sons, New Jersey, 2006.

[14] J. Borwein and S. Lewis. *Convex Analysis and Nonlinear Optimization.* Springer, New York, 2006.

[15] M. Steinbach. Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, 43:31–85, 1952.

[16] K. Falconer. *Fractal Geometry.* John Wiley & Sons, West Sussex, 2003.

[17] E. Part-Enander and A. Sjoberg. *Anvandarhandledning for MATLAB 6.5.* Uppsala University, Uppsala, 2003.

[18] M. Shishikura. The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets. *ArXiv Mathematics e-prints*, April 1992.

[19] B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. *Cowles Foundation*, 1164, 1997.

[20] L. Calvet, A. Fisher, and B. Mandelbrot. Large deviations and the distribution of price changes. *Cowles Foundation*, 1165, 1997.

[21] P. Embrechts and M. Maejima. *Selfsimilar Processes.* Princeton University Press, Princeton, 2002.

[22] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, Berlin, 1991.

[23] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes.* Chapman & Hall, London, 1994.

[24] J. Rosenthal. *A First Look at Rigorous Probability Theory.* World Scientific Publishing, Singapore, 2006.

[25] B. Mandelbrot and J. Wallis. Robustness of the rescaled range r/s in the measurement of noncyclic long-run statistical dependence. *Water Resources Research*, 5:967–988, 1969.

[26] R. Horst and N. Thoai. Dc programming: Overview. *Journal of Optimization Theory and its Applications*, 103:1–43, 1999.

[27] Nasdaq omx nordic 40 historical data. www.nasdaqomxnordic.com/aktier/historiskakurser. Accessed: 2015-08-01.

[28] Nasdaq omx nordic 40 historical index data. www.nasdaqomxnordic.com/index. Accessed: 2015-08-14.

[29] W. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.