Internal Report 2012–12 August 2012

Universiteit Leiden
Opleiding Informatica

Multiple Objective Portfolio Optimization using Evolution Strategies

Duncan Rozemond

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

Portfolio optimization is a problem that lends itself to be solved by evolutionary algorithms. This thesis is an application of three different evolutionary algorithms (SPEA2, NSGA2, SMS-EMOA) to a portfolio optimization problem using two different risk measures (variance and value at risk). Their respective fronts are evaluated using several quality indicators (error ratio, generational distance and hypervolume).

Contents

1 Introduction **6**

2 Modern Portfolio Theory **7**
2.1 Return 7
2.2 Mean-Variance 7
2.3 Mean-VaR (Value at Risk) 8

3 Multiobjective Optimization **9**
3.1 Quality Indicators 9
3.1.1 Error Ratio 10
3.1.2 Generational Distance 10
3.1.3 Hypervolume 10
3.2 Multiobjective Portfolio Optimization 11

4 Evolutionary Algorithms **13**
4.1 SPEA2 13
4.1.1 Fitness Assignment 14
4.1.2 Environmental Selection 14
4.2 NSGA-II 15
4.2.1 Fast Non-dominated Sorting 15
4.2.2 Crowding Distance 15
4.3 SMS-EMOA 17
4.3.1 S-Metric Selection 17

5 Experiments **19**
5.1 Historical Simulation 19
5.2 Experimental Setup 19
5.3 Execution 20

6 Results **21**
6.1 Mean-Variance 21
6.1.1 NSGA2 22
6.1.2 SPEA2 23
6.1.3 SMS-EMOA 23
6.1.4 Conclusion 24
6.2 Mean-VaR 25
6.2.1 NSGA2 26
6.2.2 SPEA2 26

6.2.3 SMS-EMOA 27
6.2.4 Conclusion 28
6.3 Real World Mapping 28
6.4 Conclusion 29

A Hardware and Software **31**

B Hypervolume Calculation **31**

C S&P 500 Stocks **32**

D Acknowledgements **33**

List of Figures

1 Non-dominated sorting and crowding distance used in NSGA2 courtesy of [7] . . . 17
2 Mean-Variance fronts for NSGA2, SPEA2 and SMS-EMOA . . 21
3 Mean-Variance NSGA2 front . . . 22
4 Mean-Variance SPEA2 front . . . 23
5 Mean-Variance SMS-EMOA front . . . 24
6 Mean-VaR fronts for NSGA2, SPEA2 and SMS-EMOA 25
7 Mean-VaR NSGA2 front . . . 26
8 Mean-VaR SPEA2 front . . . 27
9 Mean-VaR SMS-EMOA front . . . 28

List of Tables

1 Algorithm parameters . . . 20
2 Quality Indicators - Mean-Variance . . . 21
3 Quality Indicators - Mean-VaR . . . 25
4 Amount of V actually invested with Mean-Variance . . . 29
5 Amount of V actually invested with Mean-VaR . . . 29

List of Algorithms

1 Basic Evolutionary Algorithm . . . 13
2 fast-nondominated-sort(Q) . . . 16
3 crowding-distance-assignment($\mathcal{I}$) . . . 16
4 SMS-EMOA . . . 17
5 reduce(Q) . . . 18

Acronyms and Notation

Notation

α	Total population size
λ	Offspring population size
μ	Parent population size
μ_i	Mean (rate of) return of asset i
σ_i^2	Variance of the (rate of) return of asset i
σ_{ij}	Covariance of the (rate of) return of asset i with asset j
$\mathbb{X}$	Search (or decision) space
$\mathbb{Y}$	Objective space
$\mathcal{F}$	A Pareto front
$\mathcal{F}_{ref}$	Reference front
$\mathbf{w}$	A portfolio consisting of weights $w_1, ..., w_n$
$E(r_{\mathbf{w}})$	Expected return of a portfolio $\mathbf{w}$
G	Number of generations
m	Number of objectives
N	The total number of available assets
$\overline{N}$	Archive size
P_t	A population at time t
$r_{\mathbf{w}}$	Rate of return of a portfolio $\mathbf{w}$

Acronyms

GD	Generational Distance
IGD	Inverted Generational Distance
NSGA2	Non-dominated Sorting Algorithm 2
SMS-EMOA	S-Metric Selection Evolutionary Multiobjective Algorithm
SPEA2	Strength Pareto Evolutionary Algorithm 2

1 Introduction

Modern portfolio theory originated with the work of Harry Markowitz back in 1952. The main premise of this work was that investors need to balance the objective of maximizing the return on their investment and the minimization of the risk involved. The exact meaning of return and risk are open to interpretation. Markowitz based his model on the mean and variance of a hypothetical random variable representing future returns. Instead of using variance as a risk measure, toward the end of 1990's value at risk (VaR) became one of the most used risk measures.

This thesis applies three different evolutionary algorithms to solve the classical mean-variance problem as well as the more modern mean-VaR. The three algorithms used are SPEA2, NSGA2 and SMS-EMOA. The fronts generated by these algorithms are then examined by using several quality indicators: error ratio, generational distance and hypervolume measure.

This document is organized as follows: Section 2 will elaborate on modern portfolio theory and the risk measures used. Section 3 will look at multiobjective optimization theory and will set up the portfolio optimization problem as a multiobjective optimization problem. Section 4 will describe the evolutionary algorithms used. Section 5 will provide the details of the experiments that have been conducted and Section 6 will describe the results of those experiments.

2 Modern Portfolio Theory

Modern portfolio theory originated with the framework introduced by Markowitz in 1952 [11]. Within this framework, the target is to find an asset allocation $\mathbf{w} \in \mathbb{R}^N$ among N assets. Such an allocation should, for a given level of "expected return", minimize the "expected risk". Various approaches toward "expected return" and "expected risk" exist, of which we will discuss two: the original mean-variance approach used by Markowitz and the somewhat more modern mean-VaR (value at risk) approach.

2.1 Return

Suppose an amount X_0 is invested in an asset at time $t = 0$ and another amount X_1 is received when that asset is sold at time $t = 1$. Then the total return on the investment made in that asset is

$$R_1 = \frac{X_1}{X_0}$$

The simple term "return" is often used to refer to total return. However the same term is often also used to describe the rate of return, which is defined by:

$$r_1 = \frac{X_1 - X_0}{X_0}$$

Both concepts are related by $R_1 = 1 + r_1$ or $X_1 = (1 + r)X_0$.

More generally speaking, given a time series of prices p_t of an asset i, we can calculate the total return and rate of return over a period Δt for an asset as follows:

$$R_{t+\Delta t} = \frac{p_{t+\Delta t}}{p_t} \quad \textit{and} \quad r_{t+\Delta t} = \frac{p_{t+\Delta t} - p_t}{p_t}$$

In this paper we will only make use of the rate of return of an asset, and assume an investment period of a single day ($\Delta t = 1$) only.

2.2 Mean-Variance

The mean-variance approach is the approach initially formulated by Markowitz in 1952 [11]. This approach made use of "anticipated returns" and uses variance as a risk measure under the assumption that "an investor does (or should) consider expected return a desirable thing *and* variance of return an undesirable thing."

Mean-Variance Portfolio Given N assets with r_{it} the anticipated return at time t per unit invested in asset i, d_{it} the rate at which the return on the i^{th} asset at time t is discounted back to the present and w_i the relative amount invested in asset i. Then the discounted anticipated return of the portfolio $\mathbf{w}$ is

$$r_{\mathbf{w}} = \sum_{t=1}^{\infty}\sum_{i=1}^{N} d_{it}r_{it}w_i = \sum_{i=1}^{N} w_i \left(\sum_{t=1}^{\infty} d_{it}r_{it}\right) = \sum_{i=1}^{N} w_i r_i$$

The r_i (and $r_{\mathbf{w}}$) can be considered as being random variables. The expected return and variance of the portfolio $\mathbf{w}$ then become

$$E\left(r_{\mathbf{w}}\right) = \sum_{i=1}^{N} w_i \cdot E\left(r_i\right) \quad and \quad Var\left(r_{\mathbf{w}}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i \cdot w_j \cdot Cov\left(r_i, r_j\right)$$

where $Cov\left(r_i, r_j\right)$ represents the covariance between the returns of asset i and asset j. This dynamic model can be transformed into a static model by considering the r_i's as the "flow of returns" from the i^{th} security.

2.3 Mean-VaR (Value at Risk)

Besides variance, Value at Risk (VaR) is another widely, if not most, used risk measure. Value at Risk represents the worst expected loss over a given time horizon t at a given confidence level α and can be defined as [10]:

$$VaR_{(1-\alpha)} = -F_r^{-1}\left(\alpha\right) = -inf\left\{x \left| F_r\left(x\right) \geq \alpha\right.\right\}$$

where F_r is the cumulative distribution function of the random variable r. Here r represents profit and loss scenarios for either an individual asset or a portfolio depending on what the underlying probability distribution represents. These scenarios are calculated over the time horizon t. The confidence levels that are often used are 95% or 99%.

VaR is heavily criticized due to the fact that it does not represent a coherent risk measure[1] in the sense that diversification may actually increase the risk instead of reducing it as is expected within the framework of modern portfolio theory.

3 Multiobjective Optimization

In multiobjective (or Pareto) optimization, the purpose is to optimize a vector function $\mathbf{f} = (f_1, ..., f_m)$ consisting of m objective functions $f_i : \mathbb{X} \rightarrow \mathbb{R}$, $i \in \{1, .., m\}$ with $\mathbb{X}$ denoting the search (or decision) space. Considering only a minimization problem (without loss of generality) the problem can be formulated as follows:

$$\begin{array}{rl} \min & \mathbf{y} = f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})) \\ s.t & \mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{X} \\ & \mathbf{y} = (y_1, y_2, \ldots, y_m) \in \mathbb{Y} \end{array}$$

where $\mathbf{x}$ is called the decision vector, $\mathbf{y}$ the objective vector and $\mathbb{Y}$ the objective space.

Dominance Relation ($\prec$) Considering two objective vectors $\mathbf{y}, \mathbf{y}' \in \mathbb{Y}$, $\mathbf{y}$ is said to dominate $\mathbf{y}'$ iff

$$(\forall i \in \{1, ..., m\} : \ y_i \leq y'_i) \wedge \left(\exists j \in \{1, ..., m\} : \ y_j < y'_j\right)$$

and denote this by $\mathbf{y} \prec \mathbf{y}'$. For two decision vectors $\mathbf{x}, \mathbf{x}' \in \mathbb{X}$, we have

$$\mathbf{x} \prec \mathbf{x}' \Leftrightarrow \mathbf{f}(\mathbf{x}) \prec \mathbf{f}(\mathbf{x}')$$

Covers A decision vector $\mathbf{x}$ is said to cover another decision vector $\mathbf{x}'$ iff:

$$(\mathbf{x} \prec \mathbf{x}') \vee (\mathbf{x} = \mathbf{x}')$$

Pareto(-optimal) Set All decision vectors which are not dominated by any other decision vector of a given set are called *nondominated* regarding this set. The decision vectors that are nondominated within the entire search space $\mathbb{X}$ are denoted as Pareto optimal and constitute the Pareto-optimal front:

$$\mathbb{P} = \{\mathbf{x} \in \mathbb{X} \,|\, \nexists \mathbf{x}' \in \mathbb{X} : \ \mathbf{x}' \prec \mathbf{x}\}$$

3.1 Quality Indicators

Determining the quality of a given front is not an easy task. Although unary metrics are limited in this perspective, several metrics that may be used for this purpose are explained in [13]. Some of these metrics require a reference (or ”true") pareto $\mathcal{F}_{ref}$ representing the pareto optimal set while others don’t.

3.1.1 Error Ratio

The error ratio E_{ratio} is a ratio representing the number of points from the given front that are not members of the reference front. These members can be seen as not representing a correct solution and in that sense are erroneous. Given a front $\mathcal{F}$ consisting of n vectors s_i $(i = 1, .., n)$ the error ratio can be represented mathematically by:

$$E_{ratio} = \frac{\sum_{i=1}^{i=n} e_i}{n} \quad with \quad e_i = \begin{cases} 0 & s_i \in \mathcal{F}_{ref} \\ 1 & s_i \notin \mathcal{F}_{ref} \end{cases}$$

A value of $E_{ratio} = 0$ means that all the points of the solution front are members of the reference front and a value $E_{ratio} = 1$ means that none are.

3.1.2 Generational Distance

The generational distance (GD) is a metric representing how far a given front $\mathcal{F}$ lies from the reference front $\mathcal{F}_{ref}$. Mathematically the generational distance is defined as

$$G = \frac{\left(\sum_{i=1}^{i=n} d_i^p\right)^{1/p}}{n}$$

where n is the number of vectors in the front $\mathcal{F}$ and d_i is the distance between vector $s_i \in \mathcal{F}$ and the nearest member of the reference front $\mathcal{F}_{ref}$. If $p = 2$ then the distance metric used is the euclidean distance.

Inverted Generational Distance Instead of taking d_i the distance between $s_i \in \mathcal{F}$ and the nearest member of the reference front $\mathcal{F}_{ref}$, we can “invert" the distance by taking d_i as the distance between $s_i^* \in \mathcal{F}_{ref}$ and the nearest member of the solution front $\mathcal{F}$ and do this for each member of the reference front instead. In this case we speak of the inverted generational distance (IGD).

For both generational distance metrics, a value of $G = 0$ indicates that $\mathcal{F} = \mathcal{F}_{ref}$. Any other value indicates that the two fronts deviate from each other. The higher the value, the more they deviate.

3.1.3 Hypervolume

The hypervolume metrics (or indicator) was originally proposed by Eckart Zitzler and Lothar Thiele [15]. They referred it as “the size of the objective value space which is covered by a set of nondominated solutions". Since then it has become known as the hypervolume measure or $\mathcal{S}$-Metric and is defined as [3]:

$$\mathcal{S}\left(\mathcal{F}, \mathbf{y}_{ref}\right) = \Lambda\left(\bigcup_{\mathbf{y}\in\mathcal{F}} \left\{\mathbf{y}' \left| \mathbf{y} \prec \mathbf{y}' \prec \mathbf{y}_{ref} \right.\right\}\right)$$

where Λ is the Lebesgue measure and $\mathbf{y}_{ref}$ a reference point that is dominated by all Pareto-optimal solutions.

3.2 Multiobjective Portfolio Optimization

Multiobjective optimization when applied to the specific case of portfolio optimization comes down the following description:

Search (or Decision) Space The search space for both the Mean-Variance and Mean-VaR problems is the same. It consists of all vectors $\mathbf{w}$ (or portfolio's) in $\mathbb{X}$ such that:

$$\mathbb{X} = \left\{\mathbf{w} \in \mathbb{R}^N \left| \sum_{i=1}^{N} w_i = 1,\ 0 \leq w_i \leq 1 \right.\right\}$$

Objective Space The objective space is characterized by tuples $\mathbf{y} = (\mu_{\mathbf{w}}, \rho_{\mathbf{w}})$ where $\mu_{\mathbf{w}} : \mathbb{R}^N \rightarrow \mathbb{R}$ represents the expected return and $\rho_{\mathbf{w}} : \mathbb{R}^N \rightarrow \mathbb{R}$ represents the risk measure used. The objective space then becomes the set:

$$\mathbb{Y} = \left\{\mathbf{y} \in \mathbb{R}^2 \left| \mathbf{w} \in \mathbb{X} \right.\right\}$$

Optimization Problem The portfolio multiobjective optimization problem then becomes

$$\begin{aligned} \max \quad & \mu_{\mathbf{w}} \\ \min \quad & \rho_{\mathbf{w}} \\ s.t \quad & \mathbf{w} \in \mathbb{X}, \mathbf{y} \in \mathbb{Y} \end{aligned}$$

and is applicable to both the Mean-Variance and Mean-VaR models

Dominance . In the specific case of portfolio optimization, a portfolio with a higher return at the same (or lower) level of risk or a portfolio with a lower risk at the same (or higher) level of return are preferred. A portfolio $\mathbf{w}$ dominates another portfolio $\mathbf{w}'$ iff:

$$[(\mu_{\mathbf{w}} \geq \mu_{\mathbf{w}'}) \wedge (\rho_{\mathbf{w}} < \rho_{\mathbf{w}'})] \vee [(\mu_{\mathbf{w}} > \mu_{\mathbf{w}'}) \wedge (\rho_{\mathbf{w}} \leq \rho_{\mathbf{w}'})]$$

Covering A portfolio $\mathbf{w}$ covers another portfolio $\mathbf{w}'$ iff:

$$(\mu_{\mathbf{w}} \geq \mu_{\mathbf{w}'}) \wedge (\rho_{\mathbf{w}} \leq \rho_{\mathbf{w}'})$$

in other words if it either dominates or both have the exact same level of return and risk. The covering property of portfolios will be used extensively in the experiments (see section 5.3).

4 Evolutionary Algorithms

Evolutionary algorithms are population based stochastic global optimizers inspired by the Darwinian principle of evolution. Algorithm 1 is an outline of a basic evolutionary algorithm taken from [2].

Algorithm 1 Basic Evolutionary Algorithm

$t \leftarrow 0$
$P(t) \leftarrow initialize(\mu)$
$F(t) \leftarrow evaluate(P(t), \mu)$
while $(\iota(P(t), \Theta_\iota) \neq true)$ **do**
  $P'(t) \leftarrow recombine(P(t), \Theta_r)$
  $P''(t) \leftarrow mutate(P'(t), \Theta_m)$
  $F(t) \leftarrow evaluate(P''(t), \lambda)$
  $P(t+1) \leftarrow select(P''(t), F(t), \Theta_s)$
  $t \leftarrow t + 1$
end while

The essence of such an algorithm is the recombination-mutation-selection loop. After initialization and evaluation of an initial population P_0, this loop processes a populations P_t by using several operators such as recombination, mutation, evaluation and selection until a termination condition is reached. When the termination condition is reached, the best population and individual found during the optimization process are returned. The exact working of the algorithm depends on certain input parameters such as parent population size μ, offspring population size λ, and several others Θ_r,Θ_m, Θ_s, which are characteristic for each operator and representation of individuals. Certain EA's omit either recombination or mutation. There exist numerous variants of EA's, three of which will discussed in this section: SPEA2, NSGA2 and SMS-EMOA.

4.1 SPEA2

This section contains a brief description of the SPEA2 based on [14]

The Strength Pareto Evolutionary Algorithm 2 (SPEA2) is an improved version of the original SPEA. SPEA2 improves upon its predecessor by introducing an improved fitness scheme, a nearest neighbour density estimation technique and a new archive truncation method. SPEA2 maintains an archive $\overline{P}_t$ of fixed size ($\overline{N}$) representing the best front found so far.

4.1.1 Fitness Assignment

Each individual i in the archive $\overline{P}_t$ and the population P_t is assigned a strength value representing the number of solutions it dominates:

$$S(\mathbf{i}) = \left|\left\{\mathbf{j} \left| \mathbf{j} \in P_t + \overline{P}_t \wedge \mathbf{i} \succ \mathbf{j} \right.\right\}\right|$$

This strength value is then used to calculate the raw fitness of this individual:

$$R(\mathbf{i}) = \sum_{\mathbf{j} \in P_t + \overline{P}_t, \mathbf{j} \succ \mathbf{i}} S(\mathbf{j})$$

Besides this raw fitness value, a density estimation is performed using an adaptation of the k-nearest neighbour method. For each individual $\mathbf{i}$, the distances in objective space to all individuals $\mathbf{j}$ in the archive and population are calculated, stored in a list and sorted in increasing order. The k^{th} element in this list then represents the distance $(\sigma_{\mathbf{i}}^k)$ used to compute the density value for this individual:

$$D(\mathbf{i}) = \frac{1}{\sigma_{\mathbf{i}}^k + 2}$$

Finally, the final fitness value assigned to an individual is the sum of its raw fitness value and the density estimation:

$$F(\mathbf{i}) = R(\mathbf{i}) + D(\mathbf{i})$$

4.1.2 Environmental Selection

Environmental Selection is achieved by copying all the nondominated individuals from the current archive and population to the archive of the next generation:

$$\overline{P}_{t+1} = \left\{\mathbf{i} \left| \mathbf{i} \in P_t + \overline{P}_t \wedge F(\mathbf{i}) < 1 \right.\right\}$$

When copying these individuals, there are three cases to consider:

- $\left|\overline{P}_{t+1}\right| = \overline{N}$: The nondominated front fits exactly into the archive. In this case, environmental selection is complete.

- $\left|\overline{P}_{t+1}\right| < \overline{N}$: The archive of the next is generation is too small. In this case the best $\overline{N} - \left|\overline{P}_{t+1}\right|$ dominated individuals from the previous archive and population are copied into the new archive.

- $\left|\overline{P}_{t+1}\right| > \overline{N}$ The archive of the next generation is too large. In this case an archive truncation procedure is applied. This truncation method, iteratively removes individuals from $\overline{P}_{t+1}$ until $\left|\overline{P}_{t+1}\right| = \overline{N}$. The individual chosen for removal is the individual which has the minimum distance to another individual. If there are several such individuals, the second smallest distance is taken and so on.

4.2 NSGA-II

This section contains a brief description of the NSGA-II based on [5, 6]

The Non-dominated Sorting Genetic Algorithm II (NSGA-II or NSGA2) is an improved version of the original NSGA. This algorithm improves upon the original by introducing non-dominated sorting, crowding distance assignment and a crowded comparison operator.

4.2.1 Fast Non-dominated Sorting

The non-dominated sorting procedure (Algorithm 2) when applied to a population Q, partitions this population into v sets (referred to as fronts) $\mathcal{F}_1, ..., \mathcal{F}_v$. The first front contains all non-dominated solutions of the original population, the second contains all non-dominated solutions in the set $Q \setminus \mathcal{F}_1$, the third all non-dominated solutions in the set $(Q \setminus \mathcal{F}_1) \setminus \mathcal{F}_2$ and so on. This non-dominated sorting procedure is fast in the sense that the overall complexity is $O\left(mN^2\right)$ where N is the population size and m is the number of objectives.

4.2.2 Crowding Distance

The crowding distance estimates the density of solutions surrounding a particular solution in the population. It does so by computing the average distance of two points on either side of this solution along each of the objectives. Extremal points (i.e the solutions with the smallest and largest objective values) are assigned an infinite distance value preserving them for the next generation. The crowding distance serves as an estimate of the size of the largest cuboid enclosing the solution without including any other point of the population.

Crowded Comparison Operator Given that each individual has two attributes: a non-dominated rank (i_{rank}) and a crowding distance ($i_{distance}$), the

Algorithm 2 fast-nondominated-sort(Q)

for all $p \in Q$ **do**
  for all $q \in Q$ **do**
    if $p \prec q$ **then**
      $S_p \leftarrow S_p \cup \{q\}$
    else if $q \prec p$ **then**
      $n_p \leftarrow n_p + 1$
    end if
  end for
  if $n_p = 0$ **then**
    $\mathcal{F}_1 \leftarrow \mathcal{F}_1 \cup \{p\}$
  end if
end for
$i \leftarrow 1$
while $\mathcal{F}_i \neq \emptyset$ **do**
  $\mathcal{H} \leftarrow \emptyset$
  for all $p \in \mathcal{F}_i$ **do**
    for all $q \in S_p$ **do**
      $n_q \leftarrow n_q - 1$
      if $n_q = 0$ **then**
        $\mathcal{H} \leftarrow \mathcal{H} \cup \{q\}$
      end if
    end for
  end for
  $i \leftarrow i + 1$
  $\mathcal{F}_i \leftarrow \mathcal{H}$
end while

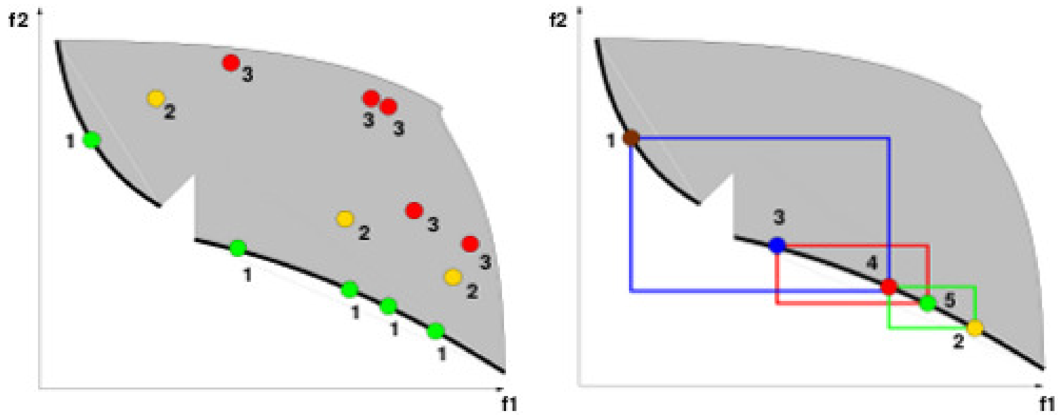
Algorithm 3 crowding-distance-assignment($\mathcal{I}$)

$l \leftarrow |\mathcal{I}|$
for all i **do**
  set $\mathcal{I}[i]_{distance} = 0$
end for
for all objective m **do**
  $\mathcal{I} = sort\,(\mathcal{I}, m)$
  $\mathcal{I}[0]_{distance} = \mathcal{I}[l]_{distance} = \infty$
  for $i \leftarrow 2$ to $l - 1$ **do**
    $\mathcal{I}[i]_{distance} = \mathcal{I}[i]_{distance} + (\mathcal{I}[i+1].m - \mathcal{I}[i-1].m)$
  end for
end for

crowded comparison operator is defined as the partial order $\prec_n$ such that:

$$i \prec_n j \Leftrightarrow (i_{rank} < j_{rank}) \vee ((i_{rank} = j_{rank}) \wedge (i_{distance} > j_{distance}))$$

In other words, the solution with the lowest rank is preferred, or when both solutions have the same rank, the one with the lowest crowding distance is preferred.



(a) Non-dominated Sorting (b) Crowding Distance

Figure 1: Non-dominated sorting and crowding distance used in NSGA2 courtesy of [7]

4.3 SMS-EMOA

This section contains a brief description of the SMS-EMOA based on [3, 8]

The S-Metric Selection Evolutionary Multiobjective Optimization Algorithm (SMS-EMOA) is an algorithm that features *fast non-dominated sorting* borrowed from NSGA-II (see section 4.2.1) and an *$\mathcal{S}$-Metric selection criterion*. Algorithm 4 is an outline of SMS-EMOA.

Algorithm 4 SMS-EMOA

$P_0 \leftarrow init()$
$t \leftarrow 0$
repeat
  $q_{t+1} \leftarrow generate(P_t)$
  $P_{t+1} \leftarrow reduce(P_t \cup q_{t+1})$ (see Section 4.3.1)
  $t \leftarrow t + 1$
until *termination condition fulfilled*

4.3.1 S-Metric Selection

After fast non-dominated sorting is applied one individual is discarded from the worst ranked front $\mathcal{F}_v$. If this front consists of more than one individual, the individual $\mathbf{i} \in \mathcal{F}_v$ that minimizes the exclusive $\mathcal{S}$-metric (see Section 3.1.3) contribution:

$$\Delta_{\mathcal{S}}(\mathbf{i}, \mathcal{F}_v) = \mathcal{S}(\mathcal{F}_v) - \mathcal{S}(\mathcal{F}_v \setminus \{\mathbf{i}\})$$

is eliminated. This measure ensures that the individuals which maximize the population's $\mathcal{S}$ metric value are kept.

Algorithm 5 reduce(Q)

1: $\{\mathcal{F}_1, \ldots, \mathcal{F}_v\} \leftarrow$ *fast-nondominated-sort*(Q)
2: $r \leftarrow argmin_{s \in \mathcal{F}_v} [\Delta_{\mathcal{S}}(s, \mathcal{F}_v)]$
3: **return** $Q \setminus \{r\}$

A steady state selection is applied due to the high computational effort of calculating the hypervolume. Only one offspring is created, requiring at most $\mu+1$ values of the $\mathcal{S}$ metric to be computed. On top of that SMS-EMOA keeps a population of non-dominated and dominated solutions in order to prevent a loss in diversity.

5 Experiments

5.1 Historical Simulation

The historical simulation is done using daily adjusted closing prices of 492 stocks taken from the S&P 500. Each of these stocks have an equal number of data points (252) over the period 2009-2010. Using these prices the rate of return of each day for each asset is calculated. These rate of returns are the basis for the historical simulation.

Mean-Variance For each asset i we can calculate the mean (μ_i) and the variance (σ_i^2) of its (rate of) return. On top of that we can also compute the covariance with any other asset j (σ_{ij}). For each portfolio $\mathbf{w}$ we can then calculate its mean return ($\mu_{\mathbf{w}}$) and variance ($\sigma_{\mathbf{w}}$).

Mean-VaR The historical simulation method used is an altered version of the one used in [4]. Given a portfolio $\mathbf{w}$ of N assets and a rate of return $r_{i,t}$ for each asset i at time t, we can construct a "scenario" for the return of the portfolio as follows:

$$R_t = \sum_{i=1}^{N} w_i \cdot r_{i,t}$$

When we do this for all t over the given period, all these scenarios form a distribution of returns over which we can calculate the VaR. Since we are using daily adjusted closing prices, the time horizon used in this historical simulation is 1 day.

5.2 Experimental Setup

Algorithm Parameters The parameters used for the algorithms are the same for both the Mean-Variance and Mean-VaR problems and are displayed in Table 1. For SMS-EMOA a child population of $\lambda = 1$ is taken due to the steady-state selection used by this algorithm. In order to obtain an equal number of function evaluations as the other algorithms, the number of generations is set to $G = 100000$.

	NSGA2	SPEA2	SMS-EMOA
Generations (G)	1000	1000	100000
Total Population (α)	100	100	100
Parent Population (μ)	100	100	100
Offspring Population (λ)	100	100	1
Crossoverrate	0.6	0.6	0.6

Table 1: Algorithm parameters

5.3 Execution

The experiment consisted of 30 runs for both Mean-Variance and Mean-VaR problem with each algorithm (NSGA2, SPEA2 and SMS-EMOA) with the given parameters. Each of these 30 runs yielded a pareto front consisting of 100 solutions resulting in a total of 3000 solutions per algorithm.

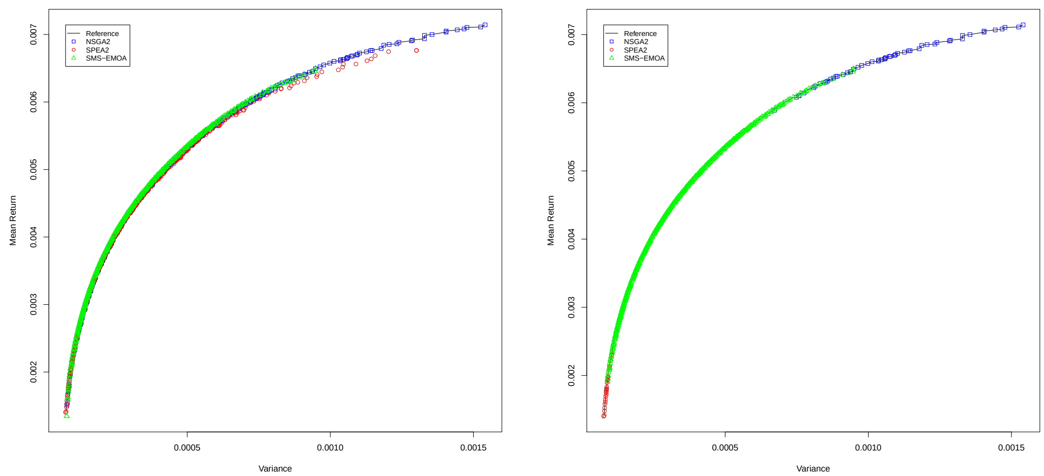
For each algorithm, from these 3000 solutions, all covering points (see section 3.2) were taken to form a final front for that algorithm. Finally, once again all covering points from all final fronts combined were taken to from a true or reference pareto front. The final front along with the reference front where then used in the analysis.

Sections A and B of the appendix explain with what software packages the experiments and analysis were performed.

6 Results

6.1 Mean-Variance

The generated fronts for the mean-variance (MV) problem can be seen in Figure 2 and corresponding quality indicators for the various fronts can be seen in Table 2. All the points combined in figure 2b form the reference front $\mathcal{F}_{ref}$.



(a) All points (b) Points on reference front

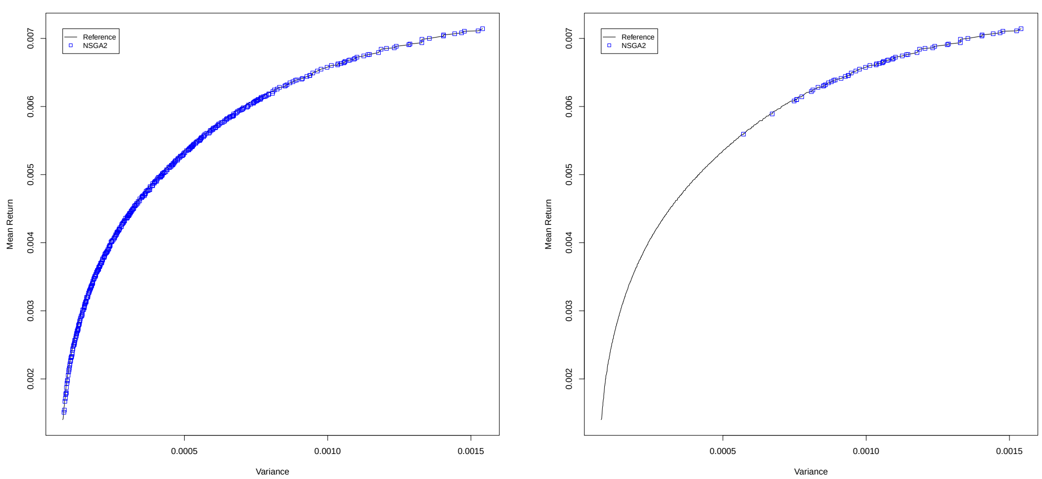
Figure 2: Mean-Variance fronts for NSGA2, SPEA2 and SMS-EMOA

	Reference Front	NSGA2	SPEA2	SMS-EMOA
Total number of points	622	435	420	**557**
Number of points on $\mathcal{F}_{ref}$	622	56	28	**538**
Contribution to $\mathcal{F}_{ref}$	100%	9.00%	4.50%	**86.50%**
Error ratio	0.00	8.712644e-01	9.333333e-01	**3.411131e-02**
GD	0.00	1.428466e-04	2.862766e-04	**3.194029e-05**
IGD	0.00	**1.259276e-04**	6.343459e-04	2.052676e-03
Hypervolume	9.741726e-01	**9.709138e-01**	9.483848e-01	9.373484e-01

Table 2: Quality Indicators - Mean-Variance

6.1.1 NSGA2

From the 30 fronts that have been generated using the NSGA2, 435 covering points have been extracted to form the final front $\mathcal{F}_{NSGA2}$ (Figure 3a). From these 435 points, only 56 contribute to the reference front (Figure 3b). These 56 points constitute approximately 9% of the reference front. The other 379 points that do not lie on the reference front create an error ratio of approximately 0.87 and generate a generational distance value of about $1.4 \cdot 10^{-4}$. For all of these indicators, NSGA2 ranks second best (or second worst). However $\mathcal{F}_{NSGA2}$ does span a hypervolume 0.97, which is the largest of the three and very close to the hypervolume of the reference front. On top of that it also has the lowest inverted generational distance of the three, with a value of approximately $1.26 \cdot 10^{-4}$.



(a) All points

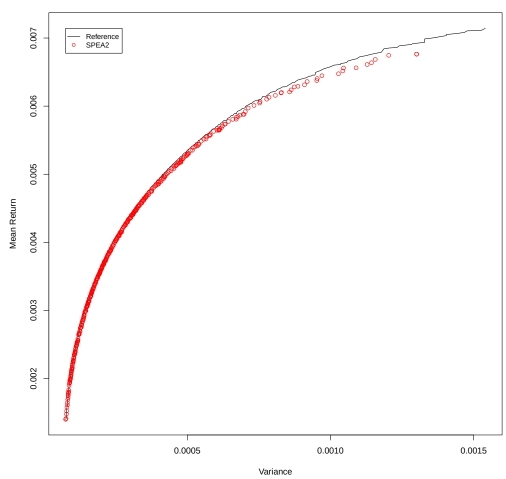
(b) Points on reference front

Figure 3: Mean-Variance NSGA2 front

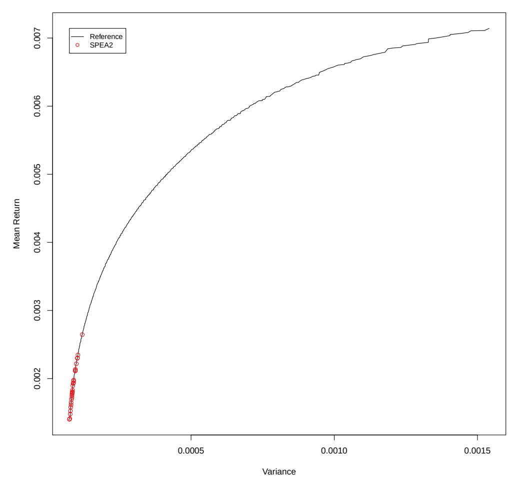
Figure 3a shows that the front generated by NSGA2 spreads out along almost the entire range of the reference front. The low inverted generational distance value show that the NSGA2 front does not lie far away from the reference front in that sense. Figure 3b shows that $\mathcal{F}_{NSGA2}$ mostly contributes to the reference front mostly in high return and high variance range, including the maximum return portfolio.

6.1.2 SPEA2

The final front for SPEA2, $\mathcal{F}_{SPEA2}$, consists of 420 points, 28 of which contribute to the reference front. These 28 points form approximately 4.5% of reference front, leading to an error ratio of approximately 0.93. The remaining 392 points account for the generational distance of approximately $2.86 \cdot 10^{-4}$. For all these indicators, SPEA2 has the worst values. However, it ranks second best (or second worst) when regarding the inverted generational distance $(6.34 \cdot 10^{-4})$ and the hypervolume (0.95) measures.



(a) All points



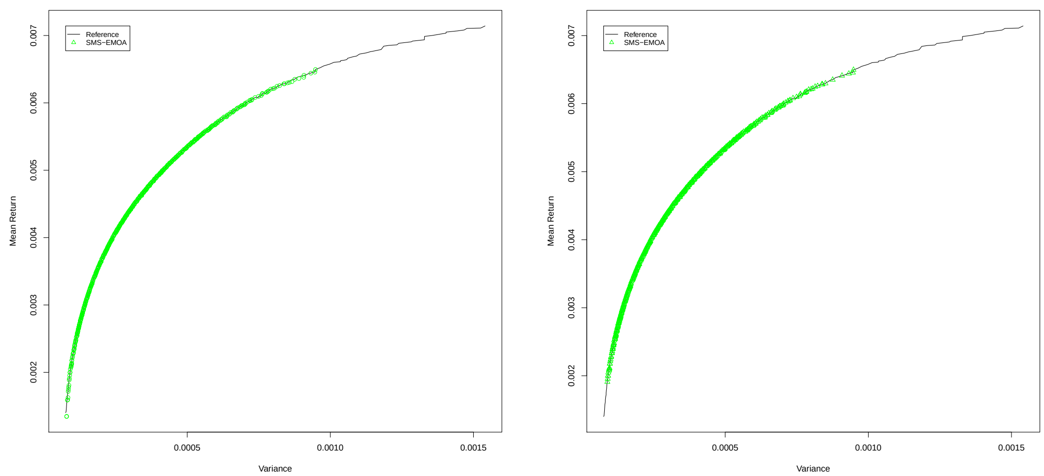
(b) Points on reference front

Figure 4: Mean-Variance SPEA2 front

Figure 4a shows that $\mathcal{F}_{SPEA2}$ deviates more from the reference the higher the variance becomes. On top of that, it does not span as big a range as NSGA2 does. All this account for the "poor" performance with regard to the quality indicators. All of the points that contribute to the reference front are in low variance low return range.

6.1.3 SMS-EMOA

The covering front generated by SMS-EMOA $\mathcal{F}_{SMS}$ has the highest total number of of points, namely 557. Out of these 557 points, 538 points contribute to the reference front. These 538 points represent 86.50% of the reference front. The other 19 generate an error ratio of approximately 0.03 and a generational distance of approximately $3.19 \cdot 10^{-5}$. For all these measures, the SMS-EMOA front comes out on top. Only on the inverted generational distance with a value of approximately $2.05 \cdot 10^{-3}$ and a hypervolume of approximately 0.94, does SMS-EMOA rank second best.



(a) All points

(b) Points on reference front

Figure 5: Mean-Variance SMS-EMOA front

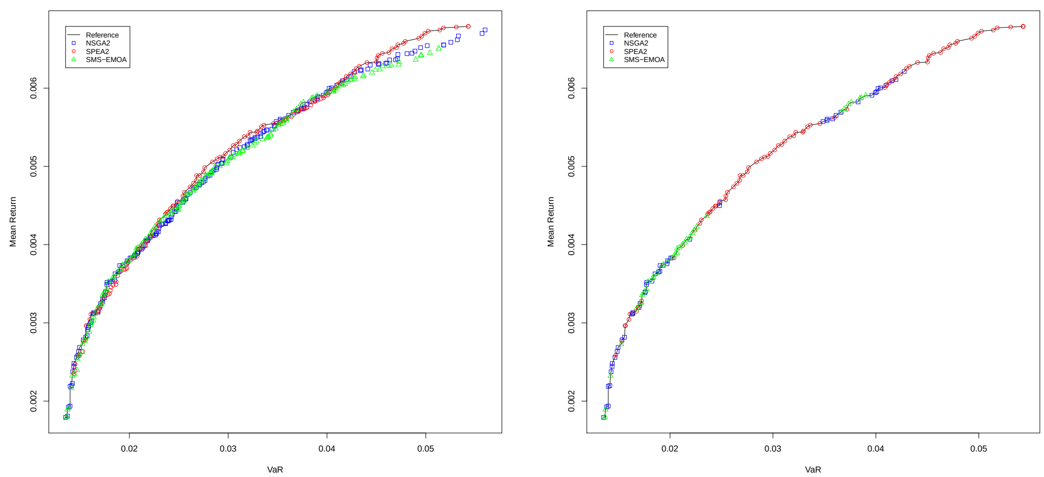
SMS-EMOA contributes by far the most points to the reference front. Figure 5a shows however that of all three fronts, $\mathcal{F}_{SMS}$ is the least spread out and spans the smallest range. This combined accounts for the low hypervolume and inverted generational distance. Since most of the points that SMS-EMOA generated are on the reference front, the generational distance and error rate are have the lowest values of the three.

6.1.4 Conclusion

From a multiobjective optimization perspective, it is not clear whether NSGA2 or SMS-EMOA is the better choice. NSGA2 is better with regard to the inverted generational distance and hypervolume measures. On top of this $\mathcal{F}_{NSGA2}$ is more spread over the objective space. SMS-EMOA however is better when considering the generational distance and error ratio. Most of the solutions generated by SMS-EMOA are non-dominated. From a portfolio optimization perspective, once again choosing an algorithm is ambiguous. Investors preferring low variance will be better off using the SPEA2 algorithm. However, as said before, from all the solutions that SMS-EMOAs generates, the greatest number of them are nondominated with respect to the other algorithms. The portfolio's that it finds are almost all among the best.

6.2 Mean-VaR

The resulting fronts for the mean-VaR problem can be seen in Figure 6a and the corresponding quality indicators can been for these fronts can been seen in Table 3.



(a) All points

(b) Points on reference front

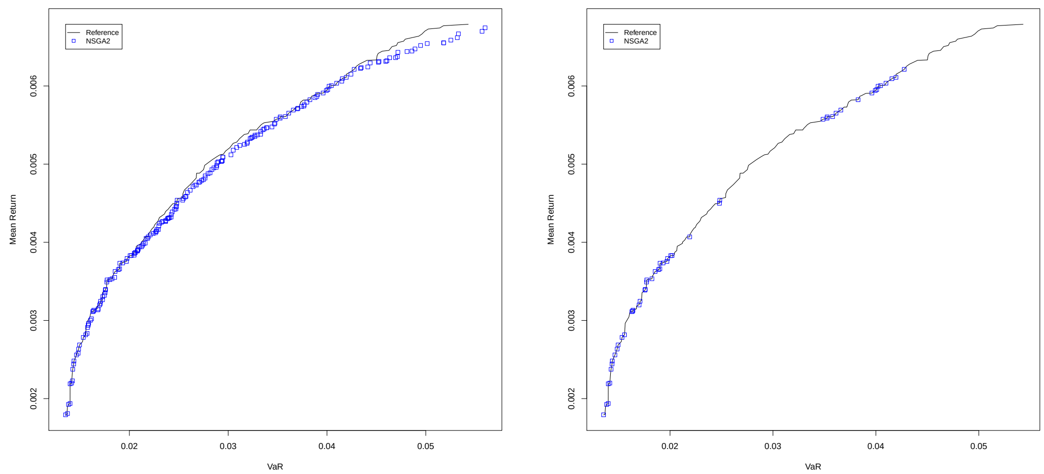
Figure 6: Mean-VaR fronts for NSGA2, SPEA2 and SMS-EMOA

	Reference Front	NSGA2	SPEA2	SMS-EMOA
Total number of points	166	**195**	137	154
Number of points on $\mathcal{F}_{ref}$	166	51	**79**	36
Contribution to $\mathcal{F}_{ref}$	100%	30.72%	**47.59%**	21.69%
Error ratio	0.0	7.384615e-01	**4.233577e-01**	7.662338e-01
GD	0.0	9.530383e-04	**4.458587e-04**	1.344539e-03
IGD	0.0	**9.262524e-04**	1.422033e-03	1.5648e-03
Hypervolume	9.010486e-01	8.854037e-01	**8.945911e-01**	8.738834e-01

Table 3: Quality Indicators - Mean-VaR

6.2.1 NSGA2

All 195 solutions generated by NSGA2 is shown in Figure 7a. Of those 195, 51 contribute to the reference front and constitute approximately 30.72% the entire reference front. The remaining 144 generate a error rate of approximately 0.74, a generational distance of $9.53 \cdot 10^{-4}$ and an inverted generational distance of $9.26 \cdot 10^{-4}$. The hypervolume measure has a value of approximately $8.85 \cdot 10^{-4}$.



(a) All points

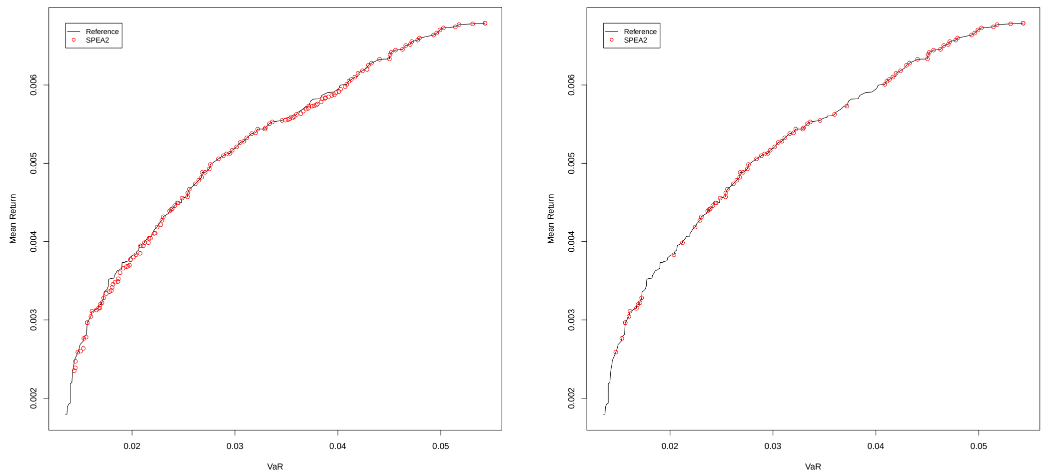
(b) Points on reference front

Figure 7: Mean-VaR NSGA2 front

The NSGA2 comes in second with respect to all but one quality indicator. It has the smallest inverted generational distance of the three. It spans along the entire range of the reference front, deviating more at the high end of the mean-variance range accounting for the best ranking when taking into account the inverted generational distance.

6.2.2 SPEA2

The final front generated by SPEA2 consists of 137 points. Of those 137 points, 79 contribute to the reference front. These 79 points represent approximately 47.59% of the reference front. The other 58 points account for an error ratio of 0.42, a generational distance of $4.46 \cdot 10^{-4}$ and an inverted generational distance of $1.42 \cdot 10^{-3}$.The entire front has a hypervolume value of approximately 0.90.



(a) All points

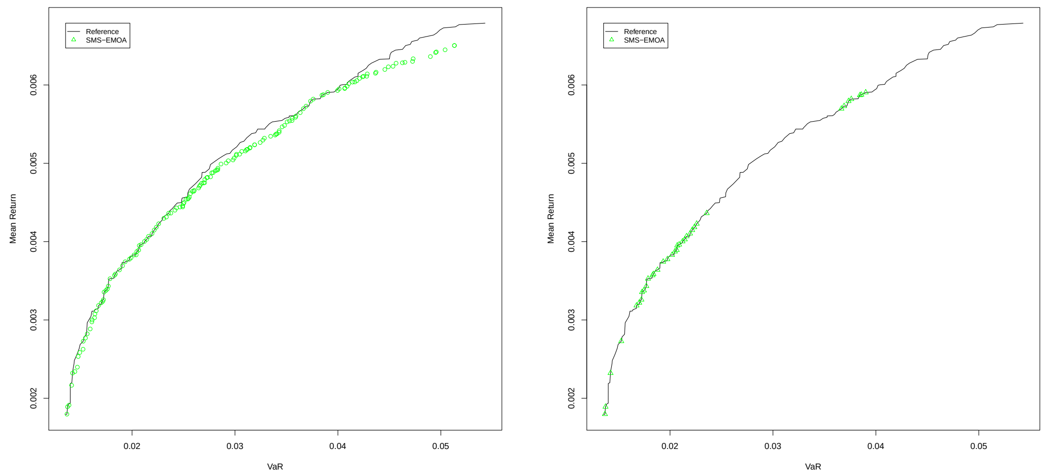
(b) Points on reference front

Figure 8: Mean-VaR SPEA2 front

SPEA2 does relatively well for the mean-VaR problem. It contributes the highest number of points to the reference front and according to most of the quality indicators, performs best. It spans a smaller range of the reference front but most of the points are either on or close to the reference point accounting for the lowest generational distance and second lowest inverted generational distance.

6.2.3 SMS-EMOA

A total of 154 points form the covering front for SMS-EMOA, 36 of which contribute to the reference front. These 36 points account for 21.69% of the reference front. The other 118 points generate an error ratio of approximately 0.77, a generational distance of approximately $1.35 \cdot 10^{-3}$ and an inverted generational distance of approximately $1.57 \cdot 10^{-3}$. The total front covers a hypervolume approximately 0.87.
SMS-EMOAs performance seems to be poor for the mean-VaR problem. It contributes the smallest number of points to the reference front. It also has the front that deviates farthest from the reference front while not spanning the entire range of this front. This accounts for both the relatively high generational and inverted generational distance measures.



(a) All points　　(b) Points on reference front

Figure 9: Mean-VaR SMS-EMOA front

6.2.4 Conclusion

The best performance on the mean-VaR problem, with regard to the performance measures, is obtained using SPEA2. Although NSGA2 is the EA finding the best minimum variance portfolio, SPEA2 does find most of the other best portfolio's. When SPEA2 does not find the best portfolio it has however found a portfolio that is close to the reference front. The worst choice however seem to be the SMS-EMOA, contributing relatively little solutions to the reference front and the ones that are not on the reference front are relatively far away from it.

6.3 Real World Mapping

Since we cannot buy fractions of a stock, we have to map the experimental result onto a real life situation. The most obvious way of doing this is to take an investment sum V and calculate the number stocks of asset i, n_i, we can really buy as follows:

$$n_i = \left\lfloor \frac{V \cdot w_i}{p_i} \right\rfloor$$

where w_i is the weight found in the experiment and p_i the adjusted closing price of asset i on the last day of the time series. Using this number we can then recalculate the weights that correspond to that number of stocks simply by:

$$w_i^* = \frac{n_i \cdot p_i}{V}$$

and once again determine the corresponding Mean-Variance and Mean-VaR fronts. Using $V = 100000$ we get the following results:

Mean-Variance

Algorithm	Mean	Standard deviation
NSGA2	0.9834920	0.001683966
SPEA2	0.9828634	0.001933367
SMS-EMOA	0.9849723	0.001650864

Table 4: Amount of V actually invested with Mean-Variance

Mean-VaR

Algorithm	Mean	Standard deviation
NSGA2	0.9834920	0.001683966
SPEA2	0.9828634	0.001933367
SMS-EMOA	0.9849723	0.001650864

Table 5: Amount of V actually invested with Mean-VaR

Overall there is not a big difference between problems nor between the algorithms and when it comes to the real world mapping. All of them generate portfolios that on average consume 98% of the total amount available. Whether it is worth the computational effort to actually try and generate portfolio that have a higher percentage invested at the same or higher levels of return or lower levels of risk remains to be seen.

6.4 Conclusion

Assessing the performance of multiobjective algorithms is difficult. As seen with the Mean-Variance, the "best" algorithm is a matter of how one looks at it, and even then the task is still difficult. What might be a good algorithm from multiobjective standpoint might not necessarily be a good algorithm when regarding it from an investor standpoint. For the Mean-VaR problem, the choice is somewhat clearer. SPEA2 does seem to “win" on most fronts.

It is obvious that the experiments conducted, have been done so using only one set of parameters on one single collection of data. To get a clearer picture, one may want to vary the parameters and underlying data more. It may very well be that given a different amount of generations or a different period over which the stock prices are considered, different results may follow.

A Hardware and Software

Software The OPT4J modular framework for meta-heuristic optimization was used for running the multiobjective optimization. The OPT4J Framework can be found at:

http://opt4j.sourceforge.net/

For the caluculation of the generational distance and inverted generational distance the quality indicator classes from the JMetal framework for multiobjective optimization were used. JMetal can be found at

http://jmetal.sourceforge.net/

For all the other analyses, the R environment for statistical computing was used. Section B in a more detailed manner in which the hypervolume computation was done.

Hardware The experiments were performed on a Pentium 4 running at 2.8GHz with 1GB Ram.

B Hypervolume Calculation

For the hypervolume calculation, the `dominated_hypervolume` from the R package 'emoa' was used. The 'emoa' R package can be found at

http://www.statistik.tu-dortmund.de/~olafm/software/emoa

This function uses the "improved dimension-sweep algorithm for the hypervolume indicator" by Fonseca et al. [9]. Version 1.3 of the hypervolume code is available from

http://iridia.ulb.ac.be/~manuel/hypervolume

The program assumes that all objectives must be minimized. On top of that, for an accurate calculation of the hypervolume measure, an normalized and positive objective space is required [8]. To this end all fronts have normalized with respect to the reference front. meaning the maximum and minimum of both objectives for the reference front have been taken to normalize all the fronts. Furthermore the mean objectives have been inverted with respect to the value 1.0, creating a pure minimization front. Finally a reference point $\mathbf{y}_{ref} = (1.1, 1.1)$ has been chosen.

C S&P 500 Stocks

MMM ACE ABT ANF ADBE AMD AES AET AFL A APD ARG AKS AKAM AA ATI AGN ALL ALTR MO AMZN AEE AEP AXP AIG AMT AMP ABC AMGN APH APC ADI AON APA AIV APOL AAPL AMAT ADM AIZ T ADSK ADP AN AZO AVB AVY AVP BHI BLL BAC BK BCR BAX BBT BDX BBBY BMS BRK-B BBY BIG BIIB HRB BMC BA BXP BSX BMY BRCM BF-B CHRW CA COG CAM CPB COF CAH KMX CCL CAT CBG CBS CELG CNP CTL CEPH CERN CF SCHW CHK CVX CB CI CINF CTAS CSCO C CTXS CLF CLX CME CMS COH KO CCE CTSH CL CMCSA CMA CSC CPWR CAG COP CNX ED STZ CEG GLW COST CVH CSX CMI CVS DHI DHR DRI DVA DF DE DELL DNR XRAY DVN DV DO DTV DFS DISCA D RRD DOV DOW DPS DTE DD DUK DNB ETFC EMN EK ETN EBAY ECL EIX EP ERTS EMC EMR ETR EOG EQT EFX EQR EL EXC EXPE EXPD ESRX XOM FDO FAST FII FDX FIS FITB FHN FSLR FE FISV FLIR FLS FLR FMC FTI F FRX FO BEN FCX FTR GME GCI GPS GD GE GIS GPC GNW GENZ GILD GS GR GT GOOG GWW HAL HOG HAR HRS HIG HAS HCP HCN HNZ HP HES HPQ HD HON HRL HSP HST HCBK HUM HBAN ITW TEG INTC ICE IBM IFF IGT IP IPG INTU ISRG IVZ IRM ITT JBL JEC JNS JDSU JNJ JCI JPM JNPR K KEY KMB KIM KG KLAC KSS KFT KR LLL LH LM LEG LEN LUK LXK LIFE LLY LTD LNC LLTC LMT L LO LOW LSI MTB M MRO MAR MMC MI MAS MEE MA MAT MKC MCD MHP MCK MWV MHS MDT WFR MRK MDP MET PCS MCHP MU MSFT MOLX TAP MON MWW MCO MS MUR MYL NBR NDAQ NOV NTAP NYT NWL NEM NWSA NEE GAS NKE NI NBL JWN NSC NTRS NOC NU NOVL NVLS NRG NUE NVDA NYX ORLY OXY ODP OMC OKE ORCL OI PCAR IR PLL PH PDCO PAYX BTU JCP PBCT POM PEP PKI PFE PCG PM PNW PXD PBI PCL PNC RL PPG PPL PX PCP PCLN PFG PG PGN PGR PLD PRU PSA PHM QLGC PWR QCOM DGX Q RSH RRC RTN RHT RF RSG RAI RHI ROK COL ROP ROST RDC R SWY SAI CRM SNDK SLE SCG SLB SNI SEE SHLD SRE SHW SIAL SPG SLM SJM SNA SO LUV SWN SE S STJ SWK SPLS SBUX HOT STT SRCL SYK SUN STI SVU SYMC SYY TROW TGT TE TLAB THC TDC TER TSO TXN TXT HSY TRV TMO TIF TWX TWC TIE TJX TMK TSS TSN TYC USB UNP UNH UPS X UTX UNM URBN VFC VLO VAR VTR VRSN VZ VIA-B V VNO VMC WMT WAG DIS WPO WM WAT WPI WLP WFC WDC WU WY WHR WFMI WMB WIN WEC WYN WYNN XEL XRX XLNX XL YHOO YUM ZMH ZION

D Acknowledgements

I would like to thank Thomas Bäck and Michael Emmerich for supervising this thesis Jetty Kleijn and Michael Lew for their time and effort with respect to the bachelorklas and my family for their support.

References

[1] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203 – 228, July 1999.

[2] Thomas Back, David B. Fogel, and Zbigniew Michalewicz, editors. *Basic Algorithms and Operators*. IOP Publishing Ltd., Bristol, UK, UK, 1st edition, 1999.

[3] Nicola Beume, Boris Naujoks, and Michael Emmerich. SMS-EMOA: Multiobjective selection based on dominated hypervolume. *European Journal of Operational Research*, 181(3):1653 – 1669, 2007.

[4] V. A. F. Dallagnol, J. van den Berg, and L. Mous. Portfolio management using value at risk: A comparison between genetic algorithms and particle swarm optimization. *Int. J. Intell. Syst.*, 24(7):766–792, 2009.

[5] Kalyanmoy Deb, Samir Agrawal, Amrit Pratap, and T Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. pages 849–858. Springer, 2000.

[6] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2000.

[7] Michael Emmerich. *Single- and Multi-objective Evolutionary Design Optimization Assisted by Gaussian Random Field Metamodels*. PhD thesis, Department of Computer Science, TU Dortmund, Germany, 2005.

[8] Michael Emmerich, Nicola Beume, and Boris Naujoks. An EMO algorithm using the hypervolume measure as selection criterion. In *2005 Int'l Conference, March 2005*, pages 62–76. Springer, 2005.

[9] C.M. Fonseca, L. Paquete, and M. López-Ibánez. An improved dimension-sweep algorithm for the hypervolume indicator. In *Evolutionary Computation, 2006. CEC 2006. IEEE Congress on*, pages 1157–1163. IEEE, 2006.

[10] Koji Inui and Masaaki Kijima. On the significance of expected shortfall as a coherent risk measure. *Journal of Banking & Finance*, 29(4):853–864, April 2005.

[11] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[12] David A. Van Veldhuizen and Gary B. Lamont. Multiobjective evolutionary algorithm research: A history and analysis, 1998.

[13] David A. Van Veldhuizen and David A. Van Veldhuizen. Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations. Technical report, Evolutionary Computation, 1999.

[14] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Technical report, 2001.

[15] Eckart Zitzler and Lothar Thiele. Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4):257–271, November 1999.