UiO : Department of Mathematics
University of Oslo

Risk - reward optimization

Illustrated with Conditional Value-at-Risk

Johannes Alnes

Master's Thesis, Spring 2016

Abstract

An important aspect in portfolio optimization is the quantification of risk. Variance was the starting point, as proposed by Harry Markowitz in the 1950's, but it's obviously flawed since it measures high returns as risk. Research have been done from both theoretical and logic point of view to improve risk measures. The result is two different groups of measures defined by axioms: (financial) deviation measures capturing the uncertainty, and risk measures which attempts to measure total exposure.

I will in this thesis present an overview of new ideas about measurement of risk. I focus especially on the work of Ralph T. Rockafellar and Stanislav Uryasev. This include Conditional Value-at-risk (CVaR), a measure that fulfills the axioms for a risk measure, and has the possibility to be solved by linear programming in an optimization model. I will also present the connected CVaR deviation. For even if risk and deviation measures are conceptual different, newer research shows an one to one relationship given certain conditions.

Lastly I compare mean-CVaR optimization with the traditional Markowitz model[1] to see if new methods actually has improved the performance of portfolio optimization. This is illustrated by optimizing the stocks in the S&P100 index for two different periods. I have also timed different solvers to show the benefits of linearization.

[1]The Markowitz model is also known as mean-variance optimization. Described in detail in chapter 3

Acknowledgements

First of I would like to thank Geir Dahl who has been my supervisor through this thesis. He took the trouble of finding me an interesting topic within the interception of finance and mathematical optimization which I enjoyed working with. He also answered every question I had with ease and in an intuitive, practical way that suited me.

I will thank my girlfriend who has supported me in times when I lacked motivation. She has surprised me with several nice dinners after extraordinary hard days at school, and taken care of everything in order for me to work hard with my thesis.

My parents have been of great support all my life and pushed me to pursue an academic degree. They also made my stay in Lausanne possible, for which I'm truly grateful.

Lastly, I will thank Hilde Sannes who with her more theoretical background has challenged me in the way I think about finance. Ekaterina Jakobsen who has made it fun being at study hall 802, and all other students that have made the last years delightful. I hope that many of the friendships from UiO will last a lifetime.

Oslo, 2016 Johannes Alnes

Contents

CHAPTER 1

Introduction

Portfolio optimization

The starting point for modern portfolio optimization was the Markowitz model by Harry Markowitz. This model was released together with the ‘’Critical line algorithm” in 1952. The critical line algorithm was the first method to find a mean-variance optimal portfolio. In 1959 Markowitz wrote the book ‘’Portfolio selection : diversification of investments”. Some areas in finance are still based on assumptions and methods from this book, like the Capital Asset Pricing Model(CAPM) and behavioral assumptions.

Many have stated Markowitz article as the starting point for ‘’modern portfolio theory”, but Markowitz have said that ‘’it’s only about portfolio theory, because it’s nothing modern about it” (Wikipedia, 2015). Even if he is most known for portfolio optimization and sparse matrix methods, Markowitz actually studied economics. While defending his PhD at University of Chicago, the concept of mathematics in portfolio theory was so new that he was criticized for presenting mathematics and not economics.

How to think about and measure risk have been studied more carefully after 1952. Markowitz himself found the biggest drawback of his model. Variance as he used as risk measure is symmetric, meaning big financial returns is counted as "risk". He realized the need for a measure capturing the idea of financial risk more closely. In 1997 Philippe Artzner wrote the article "thinking coherently" where he stated axioms for what should define a modern risk measure.

Conditional value at risk(CVaR) is one of many coherent risk measures. It’s a measure similar to value at risk (VaR)[1] which is the measure used in

[1] VaR and CVaR throughly defined in chapter 4 and 5

the Basel III and Solvency II regulations, but with more information about what happens when extreme losses occurs. In 2000 Ralph T Rockafellar and Stan Uryasev wrote "Optimization of conditional value at risk" where they showed an alternative method for optimizing CVaR. With their approach one minimizes CVaR and find VaR simultaneously, and the resulting optimization problem is on linear programming form[2].

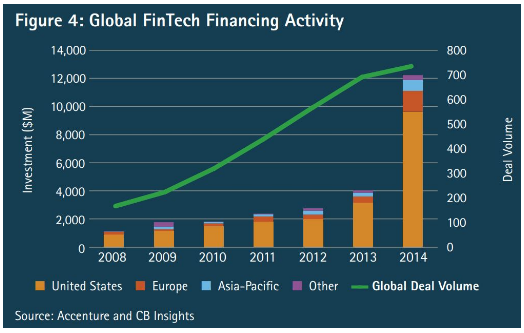
"FinTech"

FinTech is a big change going through the financial sector and stand for financial technology. As stated in (CBinsights, 2015, The Future of FinTech and Banking) : *"Global investment in financial technology or 'FinTech' spiked in 2014 reaching more than $12B in investment, demonstrating that the digital revolution in the sector is well underway."*

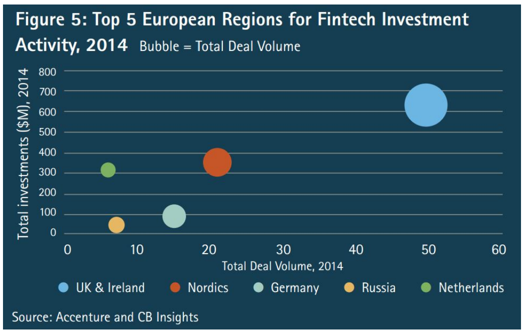
As Figure 1.1 shows, investments in FinTech have tripled in one year and the Nordic countries are a big part of it. "IKT-Norge" held many conferences in 2015 with FinTech as main point, and states Oslo has the possibility to become leading FinTech hub of Europe.

The impact of these financial changes is well illustration by Vipps. It was downloaded 100 000 times the first 10 days after launch, has changed how we transfer money and will probably change many of the online payment systems in the years to come.

[2]Optimization problem with only linear objective function and constraint. Defined in chapter 2.



(a) Global investment between 2008-2014.



(b) European investments in 2014.

Figure 1.1: Illustrations of investment done in FinTech companies. *Source: Accenture and CB insights.*

Norway has one of the most high tech bank industry in the world, where fund investing has been easy accessed on websites for a long time. Countries like the US and United Kingdom have a bank industry where fund and asset management have been expensive. It's also not an option for many people since most banks has a high minimum amount of money that is needed to invest.

Distrust and dissatisfaction about the bank industry has been a major problem after the financial crisis. Some banks invested retirement funds in assets that where so complex that risk measures were useless, hid information about where people's money were invested and gave misleading information about actual risk.

Robot advisors have therefore rapidly increased to become a big business, especially in United States. Web sites like Betterment and Wealthfront have questionnaires about your living situations, risk and investment horizon.

Based on your answers robots will give you financial advises, but will let the costumers take the final investments decisions. The costs are really low and are today at around 0.25 % of costumers investments a year in Betterment, compared to 2-3% as have been average in traditional asset management firms. This combined with minimum investment amount of 1$, absolute transparency and good looking websites and apps makes robot advisors attractive compared to the regular bank industry.

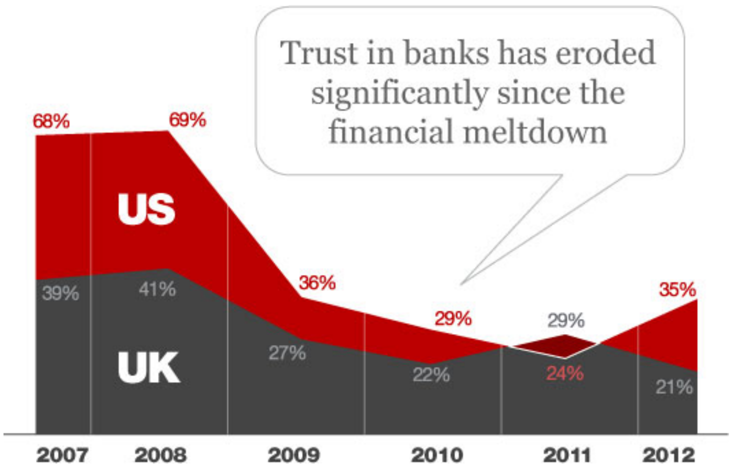


Figure 1.2: Percents of people believing banks had the capability to "do the right thing". *Copyrights: Edeleman*

Analysis of stock prices etc needs a lot of statistical methods. Optimization methods will be needed to analyze all data and conclude with a resulting asset allocation. To understand the results from statistical analysis and optimization one need some financial understanding about different assets, financial terms and assumptions used in finance. All this background information are in chapter 2.

Chapter 3 is based on the work of Markowitz. I present his mean-variance model with some of the basic assumptions for this model. From this model I defined a general risk-reward model as in "Optimal financial portfolios" by Stoyan V Stoyanov et al. I also define "the efficient frontier" which is a illustrative tool to compare risk-reward efficient portfolios.

Chapter 4 contains the axioms defined by Artzner and his logic behind coherent risk measure. Risk measures are compared to deviation measures and the connection between the two as researched by Ralph T Rockafellar

Chapter 5 takes the risk measures CVaR in the form proposed by R.T Rockafellar and Stansley Uryasev. This is shown to be a coherent measure of risk, and is presented in a mean - CVaR problem which can be solved by linear programming. I further describe different constraints that can be applied in real life optimization and different scenarios generation methods that can be used for calculations of CVaR.

Chapter 6 is numerical illustrations of portfolio optimization. I have optimized all the stocks in the US based index S&P100 for both the mean-variance model and the mean-CVaR model in order to compare performance and their efficient frontier. CVaR is dependent on the scenarios used in in the optimization problem, and four different methods are used for scenario generation throughout the results. Together these results should illustrate the financial advices that one can collect through portfolio optimization.

In chapter 7 I present the graphical user interface that I have made to this thesis. Inspired by FinTech, I made an easy to use program that downloads latest market data, gives portfolio options like risk, return and methods for solving. When optimization is done, the user can choose a portfolio on the efficient frontier, and different graphs for weight allocation, performance and histogram of VaR/CVaR together with metrics for a portfolio are shown.

CHAPTER 2

Background theory

The main goal in portfolio optimization will always be high profit with as little risk as possible. I will use $w \in \mathcal{R}^n$ as portfolio weights, and r as future returns or prices. For usual optimization we would just minimize the loss function $f(w)$, but in this setting returns are stochastic. In other words uncertainty will have to be taken into account and optimization needs to be done given probabilities for different scenarios. Following are some basic results in statistics, optimization and convexity from Devore and Berk (2007), Vanderbei (2001) and Dahl (2010), respectively, needed to understand the portfolio models. Financial facts, which are important to interpret the numerical results, are from Tsay (2005).

2.1 Statistics

Descriptive statistics

The first thing to analyze in portfolio optimization is the basic statistics: mean, variance, skewness and kurtosis. What do we expect for returns, how high is the volatility, how do different assets correspond together? All this matters and are often implemented into the optimization method.

For a continuous random variable (RV) X with a given density $p(x)$ we define the expected value as

$$E(X) = \int_{-\infty}^{\infty} xp(x)dx \tag{2.1}$$

where x is observed values of X. The n-th central moment is defined as

$$E((X - E[X])^n) = \int_{-\infty}^{\infty} (x - E[X])^n p(x) dx \tag{2.2}$$

where we have variance, skewness and kurtosis defined as the second, third and fourth moment. That is

$$\begin{aligned}
\sigma_X^2 &= E\left[(X - E[X])^2\right] \\
Skew(X) &= \frac{E\left[(X - E[X])^3\right]}{\sigma_X^3} \\
Kurt(X) &= \frac{E\left[(X - E[X])^4\right]}{\sigma_X^4}
\end{aligned}$$

Variance is one of many measures used to capture volatility in finance. Skewness and kurtosis describes the the asymmetry and tail distribution of a random variable. Last two are important when we want to assume returns follows some distribution.

In finance one have a lot of historical data that can be used to calibrate a model. Given random samples $x_1, \ldots, x_m$ from X, sample- mean, variance, skewness and kurtosis are defined:

$$\begin{aligned}
E[X] &= \frac{1}{m} \sum_{t=1}^{m} x_t \\
\sigma_X^2 &= \frac{1}{(m-1)} \sum_{t=1}^{m} (x_t - E[X])^2 \\
skew(X) &= \frac{1}{(m-1)\sigma_X^3} \sum_{t=1}^{m} (x_t - E[X])^3 \\
kurt(X) &= \frac{1}{(m-1)\sigma_X^4} \sum_{t=1}^{m} (x_t - E[X])^4 - 3
\end{aligned}$$

For mean and variance these estimations are unbiased, i.e expected value of the estimation converges to the actual value. The same does not hold for skewness and kurtosis which both is biased. Also note that I take regular kurtosis minus 3. This is called excess kurtosis, and is zero for a normal distribution.

Sample definitions taken from Tsay (2005).

Dependency

Given two independent RVs, X and Y, one are often interested in how they are related to each other. Do they change together in same direction or completely opposite of each other? The *covariance* between X and Y is defined

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= \begin{cases} \sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & \textit{if discrete} \\ \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)f(x,y)dx\ dy & \textit{if continuous} \end{cases}$$

where μ_X is expected value for X, $p(x,y)$ is the discrete joint probability mass function and $f(x,y)$ is the continuous probability density function, closer defined in the next section. When the covariance is negative the stocks move opposite of each other, zero means close to independent and they move together if positive.

In portfolio optimization the *covariance matrix* is often used, denoted Σ. Instead of two RVs, the input is a matrix of stock prices/returns $\mathbf{X} = [X_1, X_2, \ldots, X_N]$ for N different assets, and $\Sigma = cov(\mathbf{X}, \mathbf{X})$. That is, the covariance between each element in $\mathbf{X}$. The diagonal of Σ is $Cov(X_i, X_i)$ which is simply the variance of element i. As with the descriptive statistics one can calculate a sample covariance matrix from historical data

$$\Sigma = \frac{1}{(n-1)2} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i - \mu_X)(y_i - \mu_Y)$$

The downside with covariance is that the magnitude isn't comparable. That is, covariance between two RVs can be really high, but one still don't know if they are more related than two other RVs with lower covariance. Correlation, also known as standardized covariance, is another measure of dependency. Correlation always take values between -1 and 1, where negative, zero and positive have same interpretation as in covariance. This makes it easy to interpret and compare the results.

Correlation is defined

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

where σ_i is the standard deviation for RV i.

Probability density functions

A probability density function (PDF) describes the probability that X becomes a value in the interval [a,b].
Let X be a continuous random variable. Then a PDF of X is a function f(x) such that for any two numbers a,b with a≤b,

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

.
The graph of f(x) is called the density curve.

Even if it's often assumed that returns are identical and independently distributed, multivariate models are also used in finance when there is assumed some sort of relationship. Capital asset pricing model(CAPM) is one example where there is a systematic risk which influence all prices.

A multivariate distribution is described by a joint probability density function. This function describes probabilities that X takes a value in intercept of all intervals $[a_i, b_i]$, given their dependency to each other.

Let $X_1, X_2, .., X_n$ be continuous random variables. Then $f(x_1, x_2, ..x_n)$ is the joint probability density function for $X_1, X_2, .., X_n$ if for any n intervals $[a_1, b_1], [a_2, b_2], ..., [a_n, b_n]$

$$P[a_1 \leq X_1 \leq b_1, \cdots, a_n \leq X_n \leq b_n] = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, \cdots, x_n)dx_1...dx_n$$

As always with probability functions, when all limits goes to $\pm\infty$, P has to be equal to 1 and $f(x, y) \geq 0$.

The marginal distribution function for a joint distribution is given by integrating all variables but one to infinity. For instance marginal distribution for x_1 is $f_{X_1}(x_1) = f(x_1, \infty, \cdots, \infty)$. We call random variable *independent* if the joint PDF above is just the product of the independent density functions, i.e $f(x_1, \cdots, x_n) = \prod_{i=1}^n f(x_i)$.

Cumulative distribution function

The cumulative distribution function (CDF)describes the probability that X takes a value less than a limit x, and is the sum of PDF up to the point x.
The CDF, F(x), for a random variable X is defined:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)dy$$

Given that X is continuous, then at every x where derivative $F'(x)$ exists, $F'(x) = f(x)$

Similar as for PDF, the CDF for a multivariate distribution is defined

$$F_{x_1,\ldots,x_n} = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$$

Normal distribution

Normal distribution is widely used distribution in finance. Normal distribution forms the classic bell shape when plotted. The bell is centered around the mean, μ, and spread with variance, σ. The probability distribution function (PDF) is given :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp(-(x-\mu)^2/2\sigma^2) \tag{2.3}$$

where x and $\mu \in [-\infty, \infty]$ and $\sigma > 0$. Per definition a normal distribution should be perfect symmetric, i.e have skewness and excess kurtosis equal 0. We will denote a normal random variable $x \sim N(\mu, \sigma^2)$.

A normal distribution's CDF is given :

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{2\pi} exp(-t^2/2)dt, \quad -\infty < x < \infty \tag{2.4}$$

A normal distribution have two properties that are useful:

- *Invariant under linear distribution:* if $X \sim N(\mu, \sigma^2)$ and Y=aX+b, then $Y \sim N(a\mu + b, a^2\sigma^2)$
- *Linear combinations of normal distributed variables are also normal:* If $X_i \sim N(\mu_i, \sigma_i^2)$ and $Y = \sum_{i=1}^{k} a_i X_i$ then $Y \sim N(\mu, \sigma)$ where $\mu = \sum_{i=1}^{k} a_i \mu_i$ and $\sigma^2 = a_i^2 \sigma_i^2$.

Log-normal distribution

A slightly modified normal distribution is often used in finance, namely the log normal distribution. A non-negative random variable X is log-normal if $Y = ln(X)$ is normal distributed. This gives us the PDF :

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} exp(-(ln(x)-\mu)^2/2\sigma^2) & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{2.5}$$

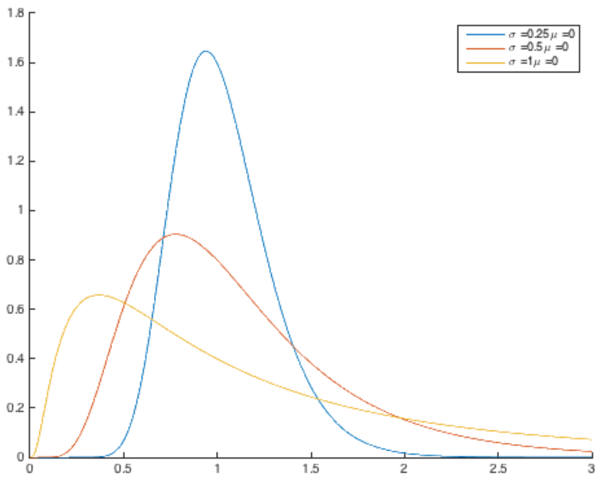


Figure 2.1: Behavior of a log-normal distribution when σ is changed. Notice how skewness changes.

It's important to notice that μ and σ is parameters for the normal distributed Y, not log-normal X. For X we got $E[X] = e^{\mu+\sigma^2/2}$ and $\sigma_X^2 = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$.

Main differences between normal and log-normal is the fact that log-normal is always non-negative. This make it useful in pricing since negative stock prices doesn't make sense. For returns on the other hand, negative values makes perfect sense and normal distribution is used.

Log-normal distribution is an example of heavy-skewed distribution. As one can see in Figure 2.1, the main mass of the distribution is centered around 0 to 1 at the x axes, but with a long right tail. This is called *positive skewness*. The skewness is controlled by σ and the distribution changes towards normal when σ goes to zero.

Student's t-distribution

When there a too few samples for a normal distribution, the result is often a Student's t-distribution. It's symmetric and bell shaped as normal distribution, but with heavier tails. Instead of μ and σ as parameters, student's t uses degrees of freedom, denoted v. When v is increasing, the student's t-distribution goes to normal. A standard student's t-distribution has PDF

$$f(t) = \frac{\Gamma(\frac{v+1}{v})}{\Gamma(\frac{v}{2})v\pi}\left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}} \quad \textit{where } \Gamma \textit{ is the gamma function.}^1$$

[1] Gamma function defined as $\Gamma(v) = \int_0^\infty t^{v-1}e^{-t}dt$

By using a scale parameter σ and mean parameter μ one can get a non-standard on the form $X = \mu + \sigma T$ where T is standard student's t-distributed. This makes it a more versatile than the normal distribution. *See Figure 2.7 in "Study of S&P" below for example.*

Empirical distribution

The empirical distribution is just a resampling of all the data that already has occurred. The empirical cumulative distribution function (ECDF) is defined as :

$$\hat{F}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{X \leq x}$$

where $\mathbb{1}_{(\cdot)}$ is the indicator function and N is number of data points. Empirical distribution is known as non-parametric distribution since one doesn't make any assumptions about distribution or parameters.

The nice thing with empirical distribution is that it will always follow the data points. This makes it good for comparison and exact when there are much data.

The drawback is that it will never take on values that haven't happened before. For instance in insurance, it will fit well until a certain percentile (90th-98th) where there are a lot of data. But there could come a natural disaster bigger than what have happened before. The empirical distribution can't forecast this. A solution could be a mixture with mainly empirical distribution, but with a heavy tailed distribution over a threshold where there are few data points.

Normal mixture distribution

Normal mixture (or Gaussian mixture) is an example of a heavy tailed distribution(high excess kurtosis). By combining two normal distributions, one can center the main mass around mean as with normal, but let one distribution has higher variance to get more extreme values. One get a mixture on the form:

$$X = (1 - L)N(\mu, \sigma_1^2) + LN(\mu, \sigma_2^2)$$

where L is a Bernoulli variable [2], such that $P(L = 1) = \alpha$ and $P(L = 0) = 1 - \alpha$. In Figure 2.2 $\mu = 0, \sigma_1 = 1$ and $\sigma_2 = 10$. What changes is the mixture probability α.

[2]

Note that one could combine more than just two normal distributions. Jump process and normal can make sense, most returns are normally distributed with big jumps representing crisis. I mention Empirical with heavy tail above, which can be mixtured in the same way. When mixed with a convex combination as above, one will get a resulting density function which integrate to 1 and is non-negative.

Drawback with mixture of distributions is that α, how much of each distribution one should combine, is hard to estimate.

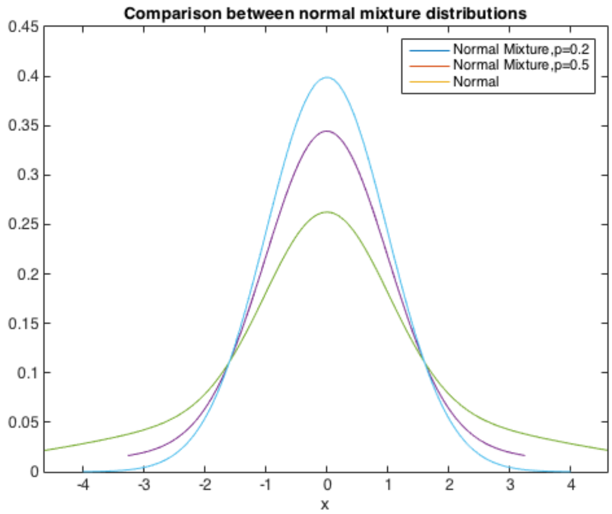


Figure 2.2: Comparison between normal mixtured. $\mu = 0, \sigma_1 = 1$ and $\sigma_2 = 10$. Same effect would happen when p is constant and σ changed.

Central limit theorem

Scenario generation is important in finance. We want to capture the effects in real world prices through simulations. One way is to simulate data by calibrating a distribution that capture the effects of interest to historical prices, and draw n numbers from this to get a sample mean $\bar{X}$.

To get higher accuracy and a confidence interval of $\bar{X}$, one can repeat this m times. If n are small $\bar{X}$ is highly influenced by outliers, but with increased n this sample mean goes to the mean from historical data. More specific:

Proposition 2.1. *Let $X_1, X_2, ..X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then:*

$$E[\bar{X}] = \mu \qquad and \qquad \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

From a normal distribution, the distribution of this sample mean will also become normal distributed and a histogram will have the bell curve.

Interestingly this is also the case for non-normal distribution like the log-normal. Sample mean tends to be more normally distributed then the original function. The Central Limit theorem states this:

Theorem 2.1. *Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then in the limit as $n \to \infty$, the standardized versions of $\bar{X}$ have the standard normal distribution. That is,*

$$\lim_{n\to\infty} P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z\right) = P(Z \leq z) = \Phi(z)$$

where $Z \sim N(0, 1)$.

This standardization can be done because of the invariant property of normal distribution stated above. This theorem states that many distributions are asymptotically normal and can be approximated by normal distribution when n is sufficiently large.

Next result that is important for sampling is the law of large numbers. It states that the probability of sample mean converges to the actual mean increases with sample size n. That is :

Theorem 2.2. *Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then $\bar{X}$ converges to μ*
I) In mean square: $E[(\bar{X} - \mu)^2] \to 0$ *as* $n \to \infty$
II) In probability : $P(|\bar{X} - \mu| \geq \epsilon) \to 0$ *as* $n \to \infty$ *for any* $\epsilon > 0$

Together this will ensure us that we can get fairly accurate results if we get scenarios based on distributions. The downside with scenarios from a distribution is that the distribution is only an assumption. This should be best fit, but could still be wrong compared to real life.

More on scenarios is discussed when we apply CVaR later on.

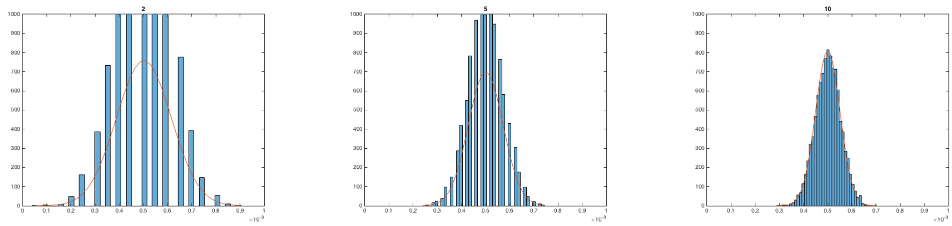


Figure 2.3: Central limit theorem applied to binomial random numbers with 10 trials and probability 0.5. Having 10 000 * i simulations where i ={2,5,10}. Clearly seeing that sample mean goes towards a normal distribution when number of simulation is increased.

Stochastic processes

For analytic finance and areas like option pricing with Black Scholes model, stochastic processes plays an important role. Fundamental in all are the Brownian motion and theory about random walk.

Brownian motion actually comes from field of particle physics, but have been generalized to a function with the following properties:

1. increments $b(t_i) - b(t_{i_1})$ is stationary
2. increments are IID - independent and identical distributed
3. $\Delta b = b(t_i) - b(t_{i-1}) = \epsilon_i \sqrt{\Delta t_i}$ where $\epsilon_i \sim N(0, 1)$

Here $t_i \in [0, T]$ and is a partitioning of a time line where b is continuous. A process with these properties is called either a standard Brownian or a wiener process. Also note that if $y = \epsilon \Delta t_i$ this imply that the increments are distributed $y \sim N(0, \Delta t_i)$.

In finance, property two tells ut that all changes are independent. What have happened before can't be used to predict the future. This is emphasized further in property three. How much a price changes is dependent on time, but in which direction is randomly drawn.

In analytical finance it is often assumed that prices have a *martingale property*, also know as the property of "fair game". This property is defined $E[r_t|r_{t-1}, ..., r_{t-t}] = r_{t-,1}$ or equivalent $E[r_t - r_{t-1}|r_{t-1}, ..., r_{t-t}] = 0$. Intuitively this says that no new information from previous event can make us predict the future any better, and best guess of tomorrows value is the present value. In financial time series it's often refereed to as " historical data has no forecasting power". See Figure 2.4 for example paths of Brownian motion with the martingale property.

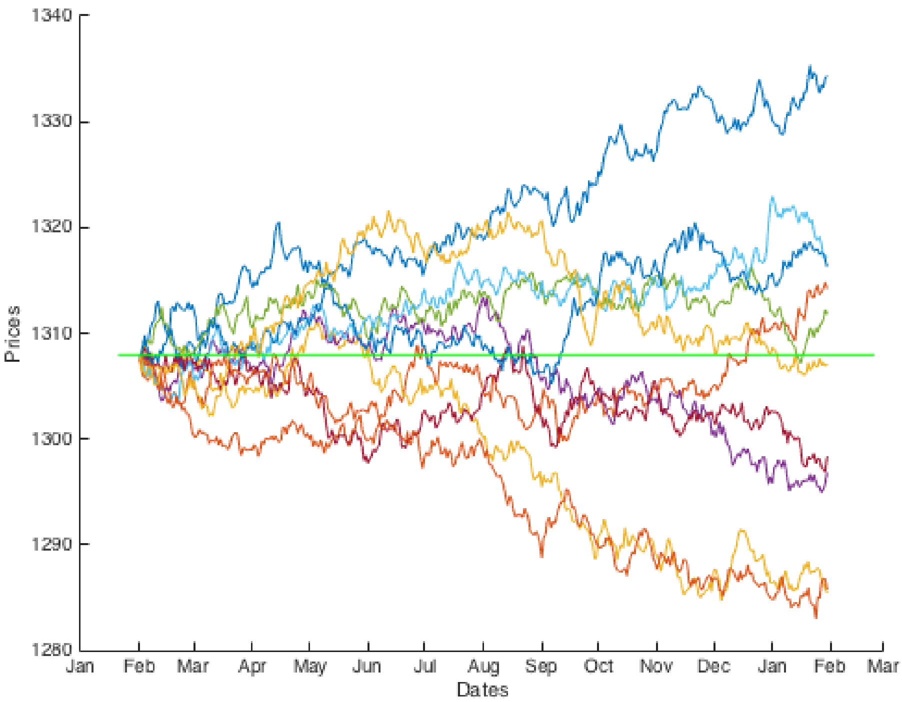


Figure 2.4: A Brownian motion as described above, calibrated to the FTSE index. Shows possible price paths for one year(251 random normal numbers)

2.2 Optimization and convexity

Convexity

Following are some basic results in the field of convexity taken from Dahl (2010).

A set $C \subseteq \mathbb{R}^n$ is called convex if for any two points, $x_1, x_2 \in C$, the point $\lambda x_1 + (1-\lambda)x_2$ also belongs to C for all $\lambda \in [0,1]$

In general we call a point $x = \sum_{j=1}^{t} x_j \lambda_j$ for $\lambda \geq 0$ and $\sum_{j=1}^{t} \lambda_j = 1$ a convex combination.

Proposition 2.2. *A set $C \subseteq \mathbb{R}^n$ is convex if and only if it contains all convex combination of it's points.*

Theorem 2.3. *Let $C \subseteq \mathbb{R}^n$ be a non-empty and line-free closed convex set. Then C is a convex hull of it's extreme points and extreme halflines, i.e*

$$C = conv(ext(C) \cup exthl(C))$$

Convex hull is defined as the set that is spanned by all possible convex combinations given by the points in that set. Theorem 2.3 states how this set will be defined, and that is a combination of all extreme points and extreme half lines.

If short-sale restrictions are added to a portfolio (all weights $w > 0$), choosing stocks are exactly such a convex combination and the set of all feasible portfolios, denoted $\mathcal{X}$, is spanned of the most extreme assets. All portfolios are a convex combination of some risky assets with high return, and some safer assets.

Let $C \subseteq \mathbb{R}^n$ be a convex set. For any two vectors $x_1, x_2 \in C$, we call a function $f : C \to \mathbb{R}$ convex if

$$f(\lambda x + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$$

Also, f is concave if -f is convex.

In general optimization one can find a minimum solution, but not be sure if it's a local or global minimum point. Given that one have a convex function, all local minimum points are also global minimum points. In other words, it's the optimal solution to the problem.

As this is a thesis about portfolio optimization some basic knowledge of different optimization models are needed. Three models are often used: Linear programming (LP), quadratic programming (QP) and convex optimization(CO). The standard formulations and models are from Vanderbei (2001).

Linear programming

Given a cost vector, $c \in \mathbb{R}^n$, we get the objective function on the linear form $c^T x$ where $x \in \mathbb{R}^n$ is the variable we want to optimize. This has to be optimized within some set defined by the matrix $A \in \mathbb{R}^{n \times n}$ and vector $b \in \mathbb{R}$. A linear program (LP) on standard form is defined :

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax \leq b \\ & x \geq 0. \end{aligned}$$

A solution x^* of this problem is a vector that satisfies the constraints and gives smaller solution, called objective value, than all other x. That is $c^T x^* \leq c^T x$ for all x. Note that minimizing $c^T x$ is equal to maximize $-c^T x$.

For LP problems we have the simplex algorithm as the preferred option. This is an old algorithm developed by G.B. Dantzig 1947 to solve the US air force's planing problems. By exploit the fact that linear constraints forms a polyhedron, the algorithms sets active number of constraints equal rank dimension for A (with added slack variables), and then "moves" through different corner solutions at the boundary to a optimal solution is found.

The simplex algorithm is not a very sophisticated, but capable of handling a lot of constraints and variables efficiently.

Quadratic programming

Quadratic programming is similar to LP, but accept a more general objective function. A QP accepts a quadratic term given by the matrix $Q \in \mathbb{R}^{n \times n}$, in addition to the linear term. On the standard form QP is defined:

$$\begin{aligned} \text{minimize } & c^T x + \frac{1}{2} x' Q x \\ \text{subject to } & Ax \geq b \\ & x \geq 0. \end{aligned}$$

Since the solution set of LP is a polyhedron, there is a unique global optimal solution. What about quadratic programming? If we assume that

Q is a positive semi-definite, then we call it a convex QP and this provides us with a unique optimum.

If not convex QP there is possibly many local minimums and one need to check each one too see which is global minimum. In the Markowitz model(minimizing variance $w^T\Sigma w$) the covariance matrix is positive semi-definite [3] This makes Markowitz model a convex QP, which is why it's easy to find an optimal solution.

Convex QP is efficiently programmed with interior-point algorithms.

Convex programming

Both LP and QP has quite strict form for the objective function and constraints. Convex optimization is more general. The objective function is now given by function $c : \mathbb{R}^n \to \mathbb{R}$ and constraints functions $g_i : \mathbb{R}^n \to \mathbb{R}$.

A convex optimization problem is on the form

$$\begin{aligned} &\text{minimize } c(x) \\ &\text{subject to } g_i(x) \leq 0 \qquad\qquad i = 1, ..., m \end{aligned}$$

Only requirements is that c is a convex function and all constraints g_i forms a convex set. As mentioned above, the set of all stocks forms a convex solutions set and will go under this category.

Since CO captures many different optimization models, one will choose solution algorithm dependent on the specific model. For instance QP and LP above are both under convex optimization, but should be solved by different algorithms.

Karush-Kuhn-Tucker theorem

The Karush-Kuhn-Tucker theorem (KKT) is often useful in non-linear optimization. This is not a model as above, but conditions a solution need to fulfill in order to be a optimal solution. The standard form of KKT is similar to CO above. Minimize $-f(x)$ subject to inequality constraints $g_i(x) \leq 0$ for $i = 1, .., m$, but in addition they add equality constraints $h_j(x) = 0$ for $j = 1, .., l$. We need to assume that the objective function $f : \mathbb{R}^n \to \mathbb{R}$, and constraints functions $g_i : \mathbb{R}^m \to \mathbb{R}$ and $h_j : \mathbb{R}^l \to \mathbb{R}$ are continuously

[3] $w^T\Sigma w = w^T E[(X - \mu_X)(X - \mu_X)]w = E[((X - \mu_X)w)^2]$. Since it's squared it's always greater or equal zero.

differentiable at point x^*. Given this KKT states four necessary conditions: stationarity, primal feasibility, dual feasibility and complementary slackness. For KKT multipliers λ_i for $i = 1, .., m$ and μ_j for $j = 1, ..., l$ the necessary conditions are:

$$\begin{aligned}
&I) \quad \nabla f(x^*) = -\sum_{i=1}^{m} \lambda_i \nabla g_i(x) - \sum_{j=1}^{l} \mu_j \nabla h_j(x^*) \\
&II) \quad g_i(x^*) \leq 0 \quad \forall i \quad and \quad h_j(x^*) = 0 \quad \forall j \\
&III) \quad \lambda_i \geq 0 \quad \forall i \\
&IV) \quad \lambda_i g_i(x^*) = 0 \quad \forall i
\end{aligned}$$

A Lagrange optimization problem is just a special case of KKT where there are no inequalities. That is the term with ∇g_i disappears and results in an easier problem. Both methods will be used in proofs of equivalence and optimality. Starting point for both are finding first order necessary conditions which are to solve property *I)*.

Source: (Krogstad, 2012)

2.3 Financial facts

Financial instruments

First of, an investor could invest in different assets:

- *Stock:* Part of a company that is out for sale, and sold to investors as shares. Gives buyer right to vote and dividends from company surplus. Only maturity is when the shares are sold or company defaults.

- *Derivatives:* More advanced contracts that gambles on the underlying movements in the markets. Often gives the right to sell or buy at certain values(options) and/or time(forward contracts). Could also be advanced contracts like credit debt swaps which gives payment if certain companies or a number of companies goes default.

- *Funds:* Putting money in certain index, sectors or investor funds. An investor with limited financial means often gets a more diversified portfolio than when picking out single equities, or one can assume that another portfolio manager, for instance at a bank, has more financial insight than yourself. Could have locked position until a maturity, meaning it's only possible to get money by selling in second hand market(real estate funds typically), or funds that one can sell at any time like stocks.

- *Bonds:* Investors can lend money to company or government, and in return get interest rates/coupon. Interest rates and often maturity (ending time) is written in the contract issued.

- *Risk free asset:* Either cash, or treasury bills issued by national banks. A treasury bill is a contract or banknote with a certain value that one can issue in the future, from 3 months to 10 years. The banknote gives no coupons or rates, but is sold for a less amount than can be withdrawn.

Portfolio allocation can be both selecting between different stocks, and choosing type of assets. Derivatives needs special calculations for each type and is not illustrated in this thesis. All other asset types has data on Yahoo and are easy to implement in the model. Treasury bills are actually auctioned, but one get more continuous data from second hand market(resale). (Wikipedia, 2016)

An investor will always have to decide two important factors. Time horizon of investment and level of risk(risk aversion). List above goes from what is risky to considered safe. One has a lot of individual risk factors for each case, which is hard to capture by a general model. For instance real estate funds. Often this is a locked position 5 to 10 years into the future, and is called *illiquid asset*. This means it's hard to sell, without loosing a lot on the trade. Index fund or stocks on the other hand is liquid, always easy to sell.

This liquidity risk are one of many factors that is specific and needs more financial understanding than what goes into an typical optimization model.

Contracts like treasury-bills and bonds has a *maturity time*. A bond would typically has three year maturity, meaning it would pay interest rates, called coupons, for three years before contract is terminated. It could be possible to sell, dependent on the contract, but bonds are more illiquid than stocks.

The main idea when allocating financial assets is *diversification*. Diversification is a theory that the total portfolio risk is lower when investments are spread on many assets, indexes and sectors.

Given all investments are in one company, and it goes bankrupt, one will loose everything. Or as we see now, oil prices is really low and the whole sector is going down. Trough diversification an portfolio would have invested in for instance renewable energy, which is performing good when non-renewable is stumbling.

A safe portfolio is composed of different asset types, different kind of risk(liquidity, volatility etc) and spread through many countries.

When buying stocks one actually invest in a specific company which in turn can give additional profit. Earnings from stocks comes by either selling at a higher price than when you bough them or getting *dividends*. Dividend is a certain amount of the companies surplus that is paid out to the shareholders of the company.

When collecting data for optimization I will use adjusted close prices. This is regular prices after stock exchange is closed, but adjusted for dividend and other similar effects.

Returns

Instead of looking directly at the price, we usually looks at the returns. A fall of 1$ could mean a big loss in one company, but with for instance Apple inc. which is worth around 110-120$ a share, it's not that drastic. Returns is change from one day to another and is easy compared between companies.

Simple return is defined $R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = (\frac{P_t}{P_{t-1}}) - 1$ where P_t is price of an asset. Simple return has the advantage that it's additive across assets. That is

$$R_t^p = \frac{\sum_{i=1}^n \pi_i P_t^i - \sum_{i=1}^n \pi_i P_{t-1}^i}{\sum_{i=1}^n \pi_i P_{t-1}^i} = \frac{1}{\sum_{i=1}^n \pi_i P_{t-1}^i} \sum_{i=1}^n \pi_i (P_t^i - P_{t-1}^i) = \sum_{i=1}^n w_i R_t^i.$$

Thus portfolio return are easily calculated. On the other side it's not additive over time, that is the geometric average doesn't equal arithmetic average $+1$.

Logarithmic return on the other hand have this property. Log return is defined as $r_t = log(1 + R_t)$ or by using prices $r_t = log(P_t/P_{t-1})$. Hence geometric average becomes $\hat{r_g} = \frac{1}{T}\sum_{t=0}^T r_t$. But for additivity across portfolio stocks

$$r_t^p = log(1 + \sum_{i=1}^n w_i(e^{r_t^i} - 1)) = log(\sum_{i=1}^n w_i e^{r_t^i})$$

which is not equal the arithmetic mean.

Study of S&P 500

S&P is short for Standard & Poor's. They have two indexes composed of the biggest 100 and 500 stocks sold on NASDAQ and NYSE, both New York based stock exchanges. Weights of the different stocks are based on market capitalization[4].

By looking at historical data from the S&P500 index(see Figure 2.5) there seems like prices grow exponential, with some trend and fluctuating variance. Returns on the other hand seems quite stable.

Stationarity is a important term in financial time series, and returns are often assumed to be weakly or covariance stationary. Weakly stationary means that the first two moments is not dependent on specific time, but only dependent on time gaps often referred to as lags. To be able to model from normal distribution for instance, weakly stationary is an important assumption. Otherwise we would need mean and variance as a function of time, and draw random samples for each time.

Some empirical fact about returns are worth noticing. First, the volatility is not time invariant. Looking at Figure 2.5 there are obviously big clustering of high volatility.

Second, there tends to be more extreme values compared to a normal distribution. And this is correct, tests like Jarque-Bera test which uses first four moments shows high kurtosis and skewness (see Table 2.1).

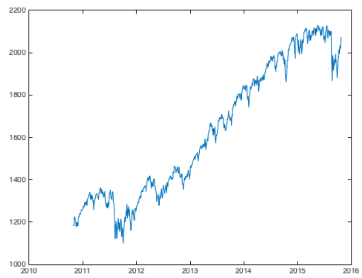
Other models are constructed to handle changing volatility, often used are ARCH/GARCH (autoregressive conditional heteroskedasticity) models which captures this cluster effects. That volatility is clustered also goes against the identical and independent distributed(IID) assumption that are the main assumption in many fields like in Black Scholes pricing.

Last, log returns are always lower than simple returns and we see they have high difference through periods with high volatility.

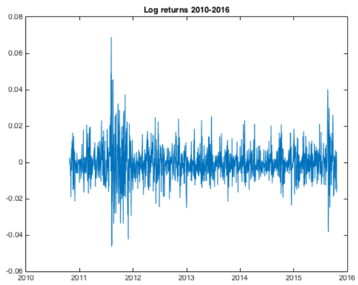
Distribution of returns

When modeling financial markets one usually assume a distribution for returns. When finding a model, time scale will affect the results. Returns for one day is something else than returns for one week. Often one will convert returns to a yearly scale for comparison, done by just multiplying mean and variance by number of time periods in a year(52 for weekly and 252 for daily).

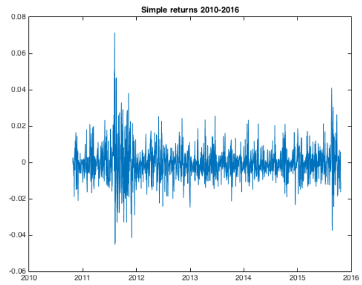
[4] *Market capitalization* is defined as number of shares available for public market times the share price. Resulting in total value of the company which is at the stock market.



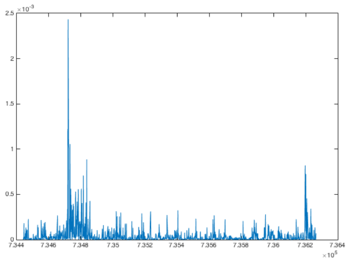
(a) Plot of price development



(b) Log returns



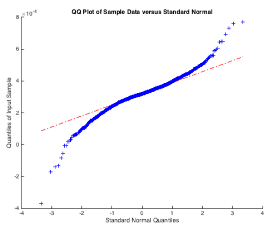
(c) Simple returns



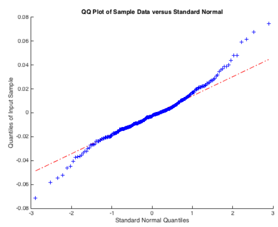
(d) Difference between simple- and log returns

Figure 2.5: S&P500 plots between 2010-2016

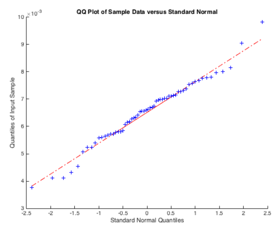
Time series facts also changes with frequency for the data. Lower frequency often gives skewness closer to zero and lower kurtosis compared to high frequency, implying that a normal distribution is often more correct with frequency at weekly or monthly data. This is illustrated through a QQ-plot, see Figure 2.6 and Table 2.1.



(a) Quantiles of daily log returns



(b) Quantiles of weekly log returns



(c) Quantiles of monthly log returns

Figure 2.6: QQplots of log returns between 2010-2016 for S&P500 index.

In Figure 2.7 normal-, student's t- and empirical distribution are fitted to weekly S&P 500 data between February 2011 to February 2016. As we

		variance	*mean*	*skew*	*kurtosis*
Simple	Daily	0.0238	-0.1020	0.6189	8.3882
	Weekly	0.0192	-0.0994	0.4195	4.9190
	Monthly	0.0140	-0.1064	0.2704	3.4587
Log	Daily	0.0237	-0.1139	0.5186	8.0912
	Weekly	0.0191	-0.1090	0.3072	4.9190
	Monthly	0.0141	-0.1138	0.1468	3.4735

Table 2.1: Descriptive statistics for simple returns and log returns based on the S&P 500 index in period 2011-2016.

see, the normal distribution with $\mu = 2.5 * 10^{-4} \approx 0$ and $\sigma = 0.01$ capture the returns quite roughly.

Better are the more heavy tailed Student-T's distribution. It's actually almost as good as the empirical distribution which follows the data almost exactly.The empirical distribution shows a higher excess kurtosis of 2.96 and skewness of 2.00, implying the data is non-normal.

Stating that returns not following normal distribution is quite drastic, since models like Black-Scholes model and CAPM, which is the most used models in financial pricing, has normal distribution as a fundamental assumptions. How much error normal assumptions is causing is a big field of study.

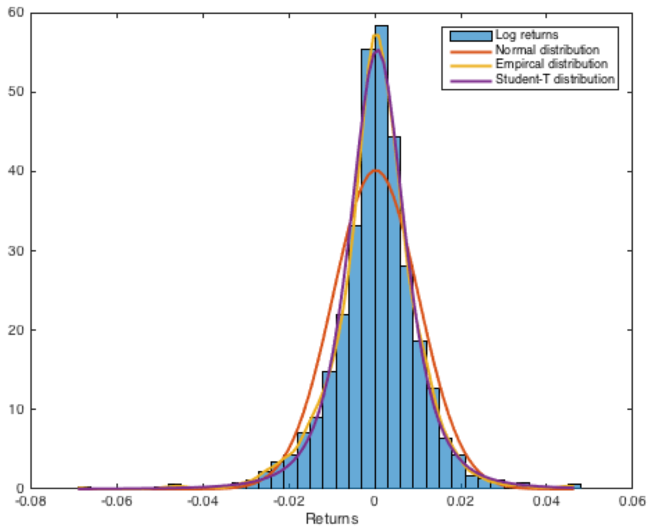


Figure 2.7: Normal, student's T and empirical distribution fitted to weekly S&P500 data between 2011-2016.

CHAPTER 3 Portfolio optimization

The difficulty in portfolio optimization is to predict how prices are going to change. In behavioral economics one will look at the underlying psychology in the markets and "bet" on certain sectors and stocks. In mathematical finance one will try to make this process objectively trough numbers. The important question I will look at in this thesis is how to think about and measure risk.

The starting point to mathematically choosing a portfolio was the Markowitz model. Harry Markowitz thought of variation in prices as the possible risk, and used variance as a risk measure. Even if more refined methods are used today, this model works well for illustrative purposes because it's quite intuitive. All ideas in this thesis have roots in the works of Markowitz, but in a more general matter.

Definition of the mean-variance model and the general risk-reward model are from Stoyanov et al. (2007). Drawbacks with mean-variance optimization are from Tsay (2005) and Investment and Financial time series courses at École polytechnique fédérale de Lausanne.

3.1 The Markowitz model

Markowitz wanted to minimize portfolio variance given a minimum level of expected return, denoted R_*. Given N assets with returns $r = \{r_1, r_2, \ldots, r_n\}$, expected return, μ_i and variance σ_i^2 for $i = 1, \ldots, n$ the Markowitz model is defined as:

$$
\begin{aligned}
\underset{w}{\text{minimize}} \quad & w^T \Sigma w \\
\text{subject to} \quad & w^T e = 1 \\
& w^T \mu \geq R_*
\end{aligned} \tag{3.1}
$$

where Σ is the $N \times N$ covariance matrix for the returns r which makes $w^T \Sigma w$ portfolio variance and $\mu_p = w^T \mu_i$ the expected portfolio return. The first constraint is to ensure an investor uses all possible capital, and the second constraint ensures return is above the given limit.

This model can also be turned around, ie maximize return given a maximum level of risk R^*.

$$
\begin{aligned}
\underset{w}{\text{maximize}} \quad & w^T \mu \\
\text{subject to} \quad & w^T e = 1 \\
& w^T \Sigma w \leq R^*
\end{aligned} \tag{3.2}
$$

Even if it's widely used the Markowitz model has some serious drawbacks.

- *Estimation errors:* sample mean and covariances are not necessarily "true" parameters. By relying on wrong parameters, resulting portfolio can be non-optimal. One solution is combining sample data with experts beliefs.
- *Sensitivity:* small changes in parameters will have a big impact on final portfolio. Large changes in position will give high transaction costs which lowers returns. Experts beliefs will stabilize the parameters, and shrinkage estimators are used to get more structured and "correct" parameters. Also robustness techniques have been applied to make the model less vulnerable to extreme values.
- *Leveraged positions:* if weights less than one is allowed, that is one short or borrows money to invest more in other assets, the model often resulting in extreme positions. The problem is, in leveraged positions the sensitivity is even higher and there is risk in borrowing money coming from need of security to get a loan that the model can't calculate.
- *Normality:* there is an underlying assumptions that returns are normal distributed, which is not the case in volatile periods

- *Behavioral:* There are assumptions about how an investor will behave. More precise, Markowitz assumed all investors should invest in one specific well diversified portfolio he called the market portfolio. This is clearly not the case in practice.

Still, the model is intuitive and one get insights in how diversifications should be done. As an analytical tool the Markowitz model is still in use, but often with newer methods for robustness and parameter estimations.

The efficient frontier

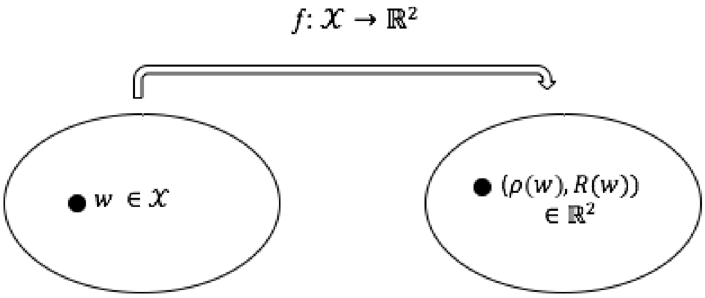
When choosing a portfolio we always optimize the different combination of assets, but the weights themselves is not what investors finds most interesting. What is more interesting is the characterization of possible financial return and risk for the portfolio, and the level of risk compared to the returns. The efficient frontier is an important tool for comparison different optimal solution within a model, as well as comparison between models.

The weight w_i is just a fraction of total value put in asset i, and for each asset there are some beliefs / calculations about risk and return. Given a market with n possible assets we get a portfolio $w = (w_1, w_2, \dots, w_n) \in \mathbb{R}^n$. Normally the possible weights will be restricted by the investors budget and other reality constraints. I will denote the set of all possible portfolio combinations as $\mathcal{X} \subset \mathbb{R}^n$.

What is of interest now is the mapping from weights to risk and reward. First off, we have a reward measure $R : \mathbb{R} \to \mathbb{R}$ and a risk measure $\rho : \mathbb{R} \to \mathbb{R}$. We can now define a function $f(w) = (x_1, x_2)$ where $x_1 = \rho(w)$ and $x_2 = R(w)$. This function is the mapping $f : \mathcal{X} \to \mathbb{R}^2$ or from the portfolio space to a "risk - reward" space. The image of this mapping will be denoted $\mathcal{S}$ and is defined by $\mathcal{S} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 = \rho(w), x_2 = R(w) \quad \textit{for some } w \in \mathcal{X}\}$.

In the Markowitz model one get $\rho(w) = \sigma_p$ and $R(w) = \mu_p$. Now $\mathcal{S}$ becomes a set in standard deviation - mean space. See Figure 3.1 for illustration. One can see from both (3.1) and (3.2) that solution set $\mathcal{X}$ is dependent on R_* or R^*. By decreasing R_* in model (3.1), that is require less return, the result is a less constrained problem. From convex theory we know this means an equal or lower objective value, which in this case means lower risk.

By increasing R_* from zero to highest possible return one get a set of different optimal portfolios $E = \{w_0(R_*) : R_*\}$. The efficient frontier is the



set $\mathcal{E} = \{(x_1, x_2) \in S : x_1 = \rho(w_0), x_2 = R(w_0)\}$ for all optimal portfolios w_0. This can also be done by changing maximum risk R^* with $E\{w_0(R^*) : R^*\}$ accordingly. Intuitively this is just setting risk to investors preferences. Look at all possible risk, and of course, the portfolio with highest rate of return at that level of risk is preferable, which is a point on the efficient frontier.

Interestingly, model (3.1) and (3.2) are equivalent in the sense that they produce the same efficient frontier. If one obtain $\sigma_p = w_0^T \Sigma w_0$ given R_* as a solution in model (3.1), one can set $R^* = \sigma_p$ and the solution of $w^T\mu$ will be R_* with portfolio weights w_0. This is proven in a general matter, after the risk-reward model is defined.

Sharpe ratio and mutual fund separation

After the portfolios on the efficient frontier is found there is the question if all portfolios are equally optimal. Robert Sharpe defined the Sharpe ratio as $SR(w) = \frac{w^T\mu}{w^T\Sigma w}$. In the Markowitz model this gives us a ratio ''expected return per unit of risk". The portfolio with highest Sharpe ratio is called tangency portfolio, since it's the point where a line from origo or a risk free rate first hits the efficient frontier.

Mutual fond separation is a theorem that is well known in finance. This theorem states that all investors will choose a combination of a common optimal risky portfolio and a risk free asset.

In the Markowitz model this states that all investors should choose a portfolio on the line from risk free rate to the tangency portfolio. This implies all investors should have the same weights on assets, but just different mix between tangency portfolio and the risk free asset. I will quickly show this analytical.

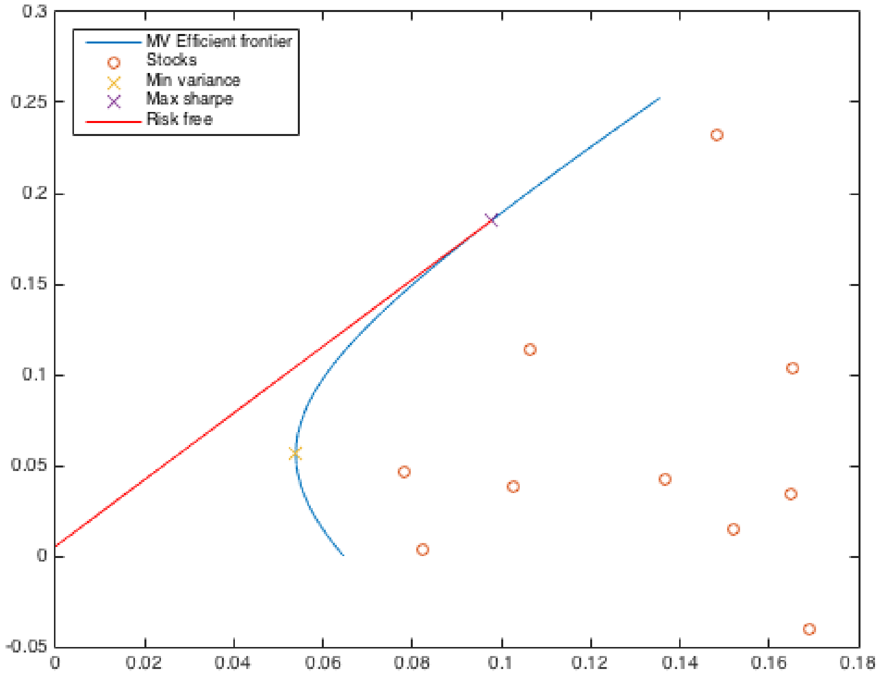


Figure 3.1: The mean-variance efficient frontier for 10 stocks with risk free option. I have marked minimum variance portfolio and tangency- /max Sharpe ratio portfolio

First we need the Lagrange for model (3.1), given by

$$L(w, \lambda) = w^T \Sigma w - \lambda(R_* - r_f - (\mu - r_f \mathbf{1})^T w)$$

.

From this we can get the first order conditions

$$\begin{aligned}
&\textit{I)} \quad \frac{\partial L}{\partial w} = \Sigma w - \lambda(\mu - r_f \mathbf{1}) = 0 \\
&\Rightarrow w = \lambda \Sigma^{-1}(\mu - r_f \mathbf{1}) \qquad (3.3) \\
&\textit{II)} \quad \frac{\partial L}{\partial \lambda} = R_* - r_f - (\mu - r_f \mathbf{1})^T w = 0
\end{aligned}$$

By manipulating *I)* one get a closed form expression for the weights. This is the essence of the mutual fond separation. w is a vector with weights of n different assets, and by chancing the constant λ, all weights change proportional to each other. λ can be seen as a risk aversion parameters, that controls mix between risk free and risky portfolio.

The amount invested in the risk free asset, w_0, is the amount not put in stocks. That is $w_0 = 1 - \mathbf{1}^T w$. The tangency portfolio on the other hand consist of risky assets only. Using this fact one get $w_0 = 1 - \mathbf{1}^T w = 0 \Rightarrow 1 = \mathbf{1}^T w$. Imposing this in the equation by multiplying *II)* with $\mathbf{1}$ gives

$$\begin{aligned} \mathbf{1}^T w =& \mathbf{1}^T \lambda \Sigma^{-1}(\mu - r_f \mathbf{1}) = 1 \\ \Rightarrow& \lambda = \frac{1}{\mathbf{1}^T \Sigma^{-1}(\mu - r_f \mathbf{1})} \\ \Rightarrow& w_{tan} = \frac{\Sigma^{-1}(\mu - r_f \mathbf{1})}{\mathbf{1}^T \Sigma^{-1}(\mu - r_f \mathbf{1})} \end{aligned} \tag{3.4}$$

where w_{tan} was found by putting λ in expression for w in (3.3). Now all investors has expected return on the form $\mu_P = \lambda w_{tan} \mu + (1 - \lambda) r_f$ dependent on risk aversion.

As mentioned Markowitz model is an good illustration, but now I want to illustrate a general model. One call equation (3.2) the mean-variance(MV) problem and is one of many possible reward - risk models with Sharpe ratio as the natural reward - risk ratio.

3.2 General reward-risk model

To ensure there exists a solution for a risk - reward model some requirements are needed. $\rho : \mathbb{R} \rightarrow \mathbb{R}$ is a risk measure which needs two properties: Positive homogeneous and sub-additive. Together these properties ensure that ρ is a convex function. The reward measure $R : \mathbb{R} \rightarrow \mathbb{R}$ needs to be a positive homogeneous and concave function. When reward and risk functions have these properties we know from the theory about convexity and optimization that there exists a solution to the reward - risk model.

The general reward - risk optimization model is defined:

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \rho(w^T r - r_f) \\ \text{subject to} \quad & w^T e = 1 \\ & R(w^T r - r_f) \geq R^* \\ & Lb \leq Aw \leq Ub \end{aligned} \tag{3.5}$$

or similar

$$
\begin{aligned}
& \underset{w}{\text{maximize}} && R(w^T r - r_f) \\
& \text{subject to} && w^T e = 1 \\
& && \rho(w^T r - r_f) \leq R_* \\
& && Lb \leq Aw \leq Ub
\end{aligned} \tag{3.6}
$$

where as before, R^* and R_* is limits for return and risk respectively.

The returns are changed to excess returns by subtracting a benchmark rate or risk free rate r_f. Putting in mean as return function and variance as risk measure gives us the original Markowitz. But we will also use measures as VaR and CVaR as ρ.

The constraints $Lb \leq Aw \leq Ub$ is general financial constraints I will illustrate later on. This could for instance limit one asset to never account for more than 20% of the portfolio value by setting $w_i \leq 0.2$. Lb and Ub are vectors in $\mathbb{R}^k$ and A some matrix in $\mathbb{R}^{n \times k}$.

As in mean-variance optimization, (3.5) and (3.6) are not equivalent in the sense that they have exactly the same problem set. But by varying minimum return and maximum risk, one get the exact same efficient frontier.

The following theorem with proof is as in (Krokhmal et al., 2002, Theorem 3, p 32). The proof sketched in this article is a bit lacking, so some supplementary details are from Krogstad (2012).

Theorem 3.1. *Let us consider risk function $\rho(w)$ and reward function $R(w)$ dependent on decision vector w, and problems:*

$$
\begin{array}{llll}
min & \rho(w) & s.t \;\; R(w) \geq R_*, & w \in \mathcal{X} \\
max & R(w) & s.t \;\; \rho(w) \leq R^*, & w \in \mathcal{X}
\end{array}
$$

Suppose the inequality constraints have internal points. Varying R_* and R^* traces the efficient frontier for the two problems. If $\rho(w)$ is convex and $R(w)$ is concave and the set $\mathcal{X}$ is a convex set, then the two problems generate the same efficient frontier.

Proof. Since this is non-linear programming we need a optimal solution to fulfill the necessary conditions defined by the Karush-Kuhn-Tucker theorem(KKT) [1]. For the problems above these conditions become:

[1]For definitions see 2.2

$$
\text{(KKT 1)} \qquad \begin{aligned} \lambda_0\rho(w^*) + \lambda_1\left(R_* - R(w^*)\right) \leq & \lambda_0\rho(w) + \lambda_1\left(R_* - R(w)\right) \\ \Rightarrow \quad \rho(w^*) - \mu_1\left(R(w^*)\right) \leq & \rho(w) - \mu_1\left(R(w)\right) \\ \mu_1\left(R(w^*) - R_*\right) = & 0, \quad \mu_1 \geq 0, \quad w \in \mathcal{X}. \end{aligned}
$$

$$
\text{(KKT 2)} \qquad \begin{aligned} -\lambda_0 R(w^*) + \lambda_1\left(\rho(w^*) - R^*\right) \leq & -\lambda_0 R(w) + \lambda_1\left(\rho(w) - R^*\right) \\ \Rightarrow \quad -R(w^*) - \mu_2\left(\rho(w^*)\right) \leq & -R(w) - \mu_2\left(\rho(w)\right) \\ \mu_2\left(\rho(w^*) - R^*\right) = & 0, \quad \mu_2 \geq 0, \quad w \in \mathcal{X}. \end{aligned}
$$

For both (KKT 1) and (KKT 2) there are three line with equation and inequalities. The first line is the necessary condition for w^* to be a optimal solution. The second line is the first line transformed to an easier form by equivalent transformation. The third line is the complementary slackness condition of KKT.

The clue for the form of KKT as written in (Krokhmal et al., 2002, Theorem 3, p 32) is that $\mathcal{X}$ is a convex set. From Krogstad (2012) there is an alternative KKT definition H.E Krogstad calls "convex KKT". Given the feasible domain is a convex set, as assumed in this theorem, the KKT necessary conditions is in addition also sufficient conditions, and an optimal solution can be proven to be an global optimal solution.

With this in mind we get that if w^* is a (global) optimal solution of model(3.5), then we know it satisfies the conditions defined by (KKT1). By setting $R^* = \rho(w)$ we see that (KKT2) is satisfied by setting $\mu_2 = \frac{1}{\mu_1}$, assuming $\mu_1 > 0$. Implying that w^* also is a solution for model (3.6), since it satisfies the necessary, and sufficient, conditions given by (KKT2).

Similarly, if w^* is a optimal solution in model (3.6) with $\mu_2 > 0$ we can set $R_* = R(w)$ in (KKT1). Now w^* is a optimal solution when $\mu_1 = \frac{1}{\mu_2}$ since all conditions in (KKT1) is satisfied. □

This proves that there are multiple way to define a portfolio optimization problem, independent of what kind of measures that is used for return and risk. But that doesn't mean they are computationally equal. Even if they finds the same solution there could be big differences speed as I will show when I compare algorithms in chapter 6.

CHAPTER 4

Risk measures

What is risk? A normal thought is that it's uncertainty connected to future prices. Then it's arguable that if an assets is guaranteed to go up with either 5$ or 10$ there will be uncertainty, but not risk since it's a sure gain. Risk is an idea of how much uncertainty there is within an asset or portfolio in combination with how exposed an investor is for losses. That is, how probable is it for an investor to get big losses.

A good measure should be able to use statistical data in an possible complex way, but also have some financial intuition. This chapter consists of axioms for deviation measures, and axioms for risk measures made to capture financial risk in a mathematical way. As it turns out, deviation and risk measures are closely related given certain properties. Both sets of definitions and their connections are from Rockafellar et al. (2006).

4.1 Deviation measures

In the following axioms some formalities are needed. As before $f(w, r)$ is a loss function which is dependent on random variable r. For the financial market there exists a large number of different scenarios for all assets. Each scenario, or state of the market, is detonated ω_i and is part of a collection of all possible scenarios denoted Ω.

Once a state has happened we get $r(\omega)$ which is constant. But the problem in portfolio optimization is the uncertainty which depending on which state the world is going to be in. When historical data are used for scenario generation, all earlier states $r(\omega_i)$ is possible with a finite number of scenarios $(\omega_1, \ldots, \omega_n)$ with related probabilities $(p_1, \ldots, p_n)$.

Trough this thesis I will defined $X = X(w, \omega) = f(w, r(\omega))$. X is therefore in the following definitions a loss. When X is defined as gain or payoff function, the axioms will look much the same, but some will have different sign.

As in Rockafellar et al. (2006) X is considered to be a random variable with a probability measure and sigma algebra, and that it's a bounded function with well defined variance and mean. Technically this gives us that X is a random variable in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$.

Definition 4.1. *A (coherent) deviation measure is a function $\mathcal{L}^2 :\rightarrow [0, \infty]$ with following properties:*

D1. *Shift invariant: $\mathcal{D}(X + C) = \mathcal{D}(X) + C$ for all X and constant C*

D2. *Positive homogeneity: $\mathcal{D}(0) = 0$ and $\mathcal{D}(\lambda X) = \lambda \mathcal{D}(X)$ for all X and all $\lambda > 0$*

D3. *Subadditivity: $\mathcal{D}(X + X') \leq \mathcal{D}(X) + \mathcal{D}(X')$ for all X, X'*

D4. *Positivity: $\mathcal{D}(X) > 0$ for all stochastically processes and $\mathcal{D}(X) = 0$ if X is constant.*

A deviation measure is how much a observed value deviate from a certain level. This could be standard deviation which measures how much observed values deviates from the mean, or median absolute deviation which measures dispersion from the median in absolute terms.

This deviation tells something about the uncertainty connected to an outcome of RV X. Both measures mentioned have a common property that is not in the list above: symmetry.

As I showed in section 2.3, returns tends to be asymmetric and heavy tailed. So how does this affect deviation measures? This means a symmetric measure doesn't necessarily capture risk in a good matter. In Markovitz's model this leads to suboptimal portfolios, since high gains isn't a ''risk", but actually preferable.

Already in 1953 Markowitz knew this and suggested a measure called ''semivariance" which didn't recognize big gains as risk, and started searching for other ways to quantify risk.

4.2 Risk measures

Risk measures and deviation measures are much alike. The conceptual difference is described above. In addition deviation measures usually capture uncertainty on the whole set, while risk measures focus on the probability for heavy downside/losses. The first is always positive, intuitively negative dispersion doesn't make sense. While risk measures looks at the probability for losses to be higher than a certain threshold value, and negative value means one can invest more while still be above threshold. This threshold could be for instance the Basel III capital requirements.

Definition 4.2. *A coherent risk measure is a function* $\mathcal{R} : \mathcal{L}^2 \to (-\infty, \infty]$ *which satisfy:*

R1. *Translation invariant:* $\mathcal{R}(X + C) = \mathcal{R}(X) + C$ *for all* X *and constant* C

R2. *Positive homogeneity:* $\mathcal{R}(0) = 0$ *and* $\mathcal{R}(\lambda x) = \lambda \mathcal{R}(X)$ *for all* X *and all* $\lambda > 0$

R3. *Subadditivity:* $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$ *for all* X, X'

R4. *Monotonicity: For all outcomes* X *and X' where* $X \geq X'$*, we have* $\mathcal{R}(X) \geq \mathcal{R}(X')$

Semivariance as Markowitz suggested is one example of risk measure defined

$$\sigma_+^2 = E[(X - E[X])_+^2] = ||(X - E[X])_+||_2^2$$

where we have $|| \cdot ||$ as the p-norm : $||X||_p = (E[|X|]^p)^{1/p}$, and $(\cdot)_+$ means only positive values i.e, $max(0, X - E[X])$.

Further we got more risk averse measures in worst case risk, widely used in robust optimization. This is defined

$$WCR(X) = supX$$

.

Both attempts to capture the asymmetry of finance, and only measures the downside of financial returns. New methods and standards are applied and we have a new group of risk measures.

The first proposal axioms for risk measures where without subadditivity. The axioms above describes a 'coherent measure of risk' according to Artzner and Delbaen (1997). Notice that since $\mathcal{R}(\lambda X + (1 - \lambda)X') \leq$

$\mathcal{R}(\lambda X) + \mathcal{R}((1-\lambda)X') = \lambda\mathcal{R}(X) + (1-\lambda)\mathcal{R}(X')$ all coherent risk measures are also convex risk measures.

In "Thinking coherently"(Artzner and Delbaen, 1997) defines these axioms from a practical point of view. Their approach is building on when a risk or position is acceptable. A risk averse investor would have less acceptance for risk, but also smaller requirements for returns than a risk loving investor.

Acceptance set is defined as the set of all positions resulting in an acceptable risk, and defined as an intersection of what is acceptable for different point of views(investor, regulator etc). The result are four axioms that defines what they all think is acceptable. One of the axioms for acceptable risk are that the acceptance set needs to be convex, which we can prove that are the case with both coherent and convex risk measures.

Artzner et al. then defined coherent risk measure trough logic based on acceptance sets. Intuitively, they interpreted the number $\rho(X)$ as the number that should be put in a risk free asset in order to make a position X to lay in the acceptance set when positive. When negative how much one can invest in risky asset to go from an acceptable to unacceptable position. In other words, ρ is in many ways a measure of capital requirement which needs to be satisfied if one should invest in the portfolio, and is acceptable when $\rho \leq 0$.

In this sense the axioms above means:

- *Translation invariant* consider the case where a total future value of $X - m$ is given, where m is certain payoff, but X is stochastic loss. The total capital requirement is decreased with m since it doesn't hold risk and can be set outside of the risk function.

- *Positive homogeneity* states that if one increases a position in X the risk is increased proportionally.

- *Subadditivity*y tells us that investing in different assets will result in equal or less risk. Or as (Artzner and Delbaen, 1997) states: "*a merger does not create extra risk*".

- *Monotonicity* tells that if one have two future values (random variables) where it's know that one outcome has higher loss than the other, $X \geq X'$, then the required capital will be higher for this random variable, $\mathcal{R}(X) \geq \mathcal{R}(X')$. This because it has higher downside.

- *Convexity* describes the diversification effect. Instead of risking everything in X or Y, one take a fraction (λ) in each. By investing everything in X one get $\rho(X)$ and similar for Y. Thus a combination of both will give equal or less risk by definition of convexity.

4.3 The connection between risk and deviation

As we see the axioms for deviation and coherent risk measures are quite similar, and research have tried to establish a connection between the two. Risk measures have the benefit of better financial intuition and made explicit for risk. The consequence is that a risk measure usually only uses a part of the dataset which increases error.

In Rockafellar et al. (2006) a fifth axiom was added to the definition of coherent risk:

$$\textbf{R5.}: \quad \mathcal{R}(X) > E[-X] \quad \textit{for all nonconstant } X.$$

called strict expectation boundness.

Rockafellar et al. showed that when a risk measure fulfilled **R1-R3** and **R5** they had a one to one connection to deviation measures given:

$$\mathcal{R}(X) = \mathcal{D}(X) - E[X] \tag{4.1}$$

$$\mathcal{D}(X) = \mathcal{R}(X - EX) \tag{4.2}$$

If risk measure $\mathcal{R}(X)$ is found trough 4.1, then it is a coherent measure of risk if and only if $\mathcal{D}(X)$ is upper range dominated. *Upper range dominated* mean that the deviation always has to be smaller than the difference between the mean of position X and the maximum connected to position X, $\mathcal{D}(X) < supX - E[X]$ *for all X.* In other words, highest loss is higher than highest return which often is the case in the financial market.

In practice we get for instance that standard deviation, $||X - E[X]||_2$ will be connected to a "standard deviation risk" measure defined as $||X - E[X]||_2 - E[X]$, and with more refined risk measures there will also follow more capable deviation measures[1].

4.4 Value at risk and Conditional value at risk

Definition 4.3. *Given a stochastic variable r with decision vector w, we want to find the smallest number ζ such that the probability that loss $X = f(w, r)$*

[1]For proofs see Rockafellar et al. (2006) and for more about risk and deviation measures seeKrokhmal et al. (2011)

exceeds ζ is no more than (1-α) where $\alpha \in (0,1)$

$$VaR_\alpha(r) = inf(\zeta \in \mathbb{R} : P(X \geq \zeta) > 1 - \alpha)$$

Note that X is a change of value over a certain time period. This could for instance be the loss over a 10 day period which is $f(w,r) = V_{t+10} - V_t$, where V_t is total value of the portfolio. If calculation of VaR gives 100 000$ for $\alpha = 0.95$ and loss function as described, this intuitively means that losses above 100 000$ over a 10 day period could happen once every twentieth period.

A monetary value is often used when describing VaR, but loss function with returns instead of prices, like $f(w,r) = \frac{v_{t+10}}{v_t} - 1$ is also used. VaR is then a percentage of the portfolio value that can be lost with probability α, and one can multiply VaR with total value to get a monetary value.

Since VaR in many cases lacks sub-additivity, new measures have been proposed to take it's place. Especially conditional value at risk have nice features.

Definition 4.4. *For a loss X and a confidence level α we define conditional value at risk(CVaR) as*

$$\begin{aligned} CVaR(r)_\alpha &= E[X | X \geq VaR_\alpha] \\ &= \frac{1}{1-\alpha} \int_\alpha^1 VaR_u(r) du \end{aligned}$$

Expected shortfall, average VaR and expected tail loss is describing the same thing as CVaR, but defined in slightly different ways. All measures see what values to expect given that the loss is greater than VaR. When distribution function of r is continuous, as assumed here, they're all equivalent.

4.5 Comparison

VaR

- Industry standard in finance and is used as requirement in Basel. Much because of it's ease of use and intuitiveness.
- Can be estimated parametric when a distribution is assumed.
- More numerical stable to outliers than CVaR. Extreme event tends to be hard to measure and therefore gives higher errors. VaR discard many of these values.

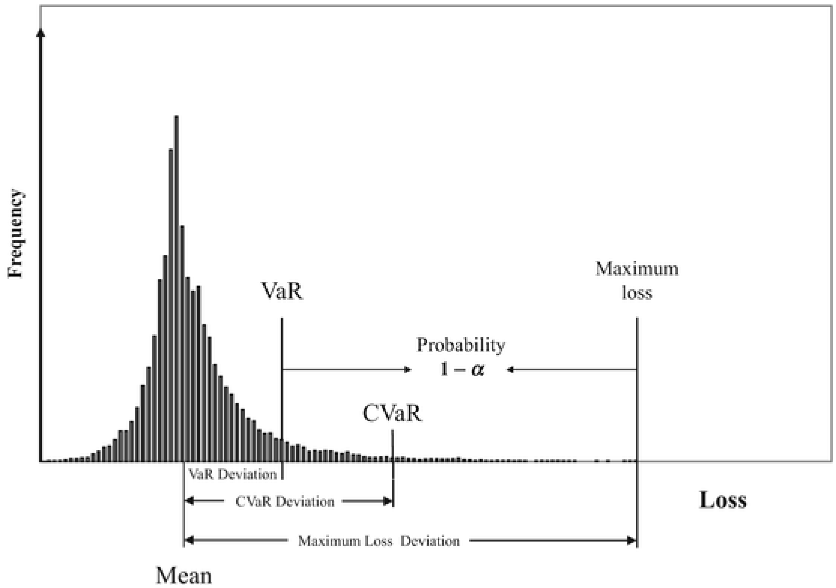


Figure 4.1: Illustration of VaR and CVaR on histogram of returns. Note that $VaR \leq CVaR$.

- *Not a coherent measure of risk, lacks sub-additivity.*
- *When there is correct data in tail and there is many extreme events, VaR can be misleading since the extreme events aren't used in calculation.*
- *Since VaR lack sub-additivity it's not convex and therefore harder to use in optimization problems*
- *Less stable than CVaR when α is changed. Small changes in α can leads to big jumps, since all information in the tail is discarded.*

CVaR

- A coherent risk measure, implying it's easy to use in optimization problems.
- Measuring all heavy losses, and optimization with CVaR as risk measure therefore has more control of extreme losses than VaR
- Can be linearized when there are generated scenarios.
- *Often lacking data in the tails can imply unreliable models.*

- *Little data and the us of all extreme values leads to higher sensitivity to estimation errors than VaR.*

VaR is today industry standard in both finance and insurance, mostly because it's easy to understand and compute. Mathematically and in an optimization model it's not that convenient. Basel have been criticized for using VaR as standard, and Embrechts et al. (2014) goes as far as saying ''Value-at-Risk can destabilize an economy and induce crashes".

Most criticism is connected to the fact that it's not sub additive and thus not a coherent risk measure. This is the property that states that one can diversify and thus get smaller risk by having a portfolio of many asset from different classes. That VaR doesn't capture sub additivity is easily shown by a jump process.

Jump processes often happens for derivatives that gambles on if a company defaults or not. Given two equal bonds with a small probability, for instance 3 %, for default which comes with a big loss or zero otherwise. For both bonds the 95% VaR is zero, since the probability for default are lower than 5%. Given these default happen independently, the chance for both going default is 0.1%, that one defaults is 5.82%, and none 94%. But then the VaR is increased to a big loss and not zero as before when combined because of the increased probability for one bond going default.

More relevant for this thesis is the fact it's not convex (convexity shown by using sub additivity above) and behave badly in an optimization problem. As we know convexity gives an unique optimal solution, and by lacking convexity optimization of VaR can give multiple optimums.

There are cases where VaR is coherent, and that is when returns follows an elliptical distribution. Normal distribution is one example of elliptical distribution, and it can be shown that in this case mean - VaR analysis yields the same results as with mean - variance analysis. But then it's not necessary to use the computational harder mean - VaR model. [2]

There are many features that makes CVaR a better risk measure than VaR. From a financial point of view it gives more secure decisions. VaR don't say anything about what happens given a extreme loss, but CVaR will in such case give a number of how much a firm can loose, taken all information into account. From a mathematical point of view CVaR is convex for given level α which makes it easy to optimize.

[2] For proof and more details of elliptical distributions see Landsman and Valdez (2003) and easy proof for normal distribution see Rockafellar and Uryasev (2000).

The methods used for comparing VaR and CVaR are criticized. Many compare VaR and CVaR at a given level α, but they often are at two different places in the distribution(for illustration see 4.1). In general one have $VaR_\alpha \leq CVaR_\alpha$, and thus VaR is in a more stable part of the distribution. But one can always find α_1 such that $CVaR_\alpha = VaR_{\alpha_1}$.

Some criticism and results are found in Uryasev et al. (2010). Clearly the longer out in the tail a threshold gets, the more unstable it is. Uryasev et al found cases where $CVaR_{0.95} = VaR_{0.99}$ which is in a more unstable region than $VaR_{0.95}$.

As a conclusion between the two, and especially with stability in mind they stated *"Thus, one should analyze properties of the dataset on which computations are based, with particular focus on the model for the tails of the distribution, before deciding to insert constraints on VaR or CVaR, as none of them is "better" than the other"*. Which is interesting since many sees CVaR as superior in terms of mathematical properties, but evidently not always in practice.

4.6 VaR- and CVaR deviation

Ralph T. Rockafellar have been behind much of the research of CVaR, the study of connections between risk and deviation measures, and is behind the one to one connection.

He have also developed CVaR deviation which he describes as promising. This because he in Rockafellar et al. (2006) shows it can directly replace standard deviation as deviation measure in most of the concepts developed by Markowitz like mutual fond separation, the Markowitz model and the Sharpe ratio.

Remember the connection (4.2) and with VaR and CVaR as risk measures we get the associated deviation measures :

Definition 4.5.

$$\begin{aligned}
\mathcal{R}(X) = VaR_\alpha(X) \quad &\longleftrightarrow \quad \mathcal{D}(X) = VaR_\alpha(X - E[X]) \\
\mathcal{R}(X) = CVaR_\alpha(X) \quad &\longleftrightarrow \quad \mathcal{D}(X) = CVaR_\alpha(X - E[X]) = CVaR_\alpha^{\triangle}(X)
\end{aligned}$$

Notice that CVaR is a coherent risk measure and is strictly expectation bounded, implying that the CVaR deviation is a coherent deviation measure. VaR on the other hand fails on sub-additivity and monotonicity in certain cases. Meaning the VaR deviation is not fulfilling the axiomatic definition of a deviation measure.

The CVaR deviation has become increasingly of interest because of new methods connected to this measure. Sarykalin et al. (2008) states that "*Another coherent deviation measure in the basic sense is the so-called Mixed Deviation CVaR, which we think is the most promising for risk management purposes.*"

Mixed CVaR deviation is a convex combination of CVaR where different levels of α gets weighted. That is

$$Mixed - CVaR^{\triangle} = \int_0^1 CVaR_{\alpha}^{\triangle}(X) d\lambda(\alpha)$$

where λ *is a weighting measure on* $(0,1)$ *and total measure 1.* Now, instead of just a risk parameter λ one can define a total risk profile by weighting λ different for each α. Making it possible for more complex and structured models of optimization.

Each different deviation and risk measure described in this chapter are possible to use as ρ in the reward - risk model, but with the drawbacks described for non coherent measures.

CHAPTER 5

Mean - CVaR optimization model

Now I want to illustrate how reward-risk analysis could look like beside Markowitz. As before, $r \in \mathbb{R}^n$ equals returns and $w \in \mathcal{X}$ is the portfolio weights with loss function $f(w, r)$.

Formulations connected to measures and probability is on the same form as in Krokhmal et al. (2002). For convenience I will assume returns have a probability distribution, defined as $p(r)$, but this can be avoided. See (Rockafellar and Uryasev, 2002).

The cumulative distribution function for $p(r)$ can be used to set 'maximal acceptable level of loss'. Given a portfolio with weights w we can change the loss threshold, ζ, to get a cumulative distribution function

$$\Psi(w, \zeta) = \int_{f(w,r)\leq\zeta} p(r)dr. \tag{5.1}$$

where $\Psi(w, \zeta)$ is continuous because of the assumptions of $p(r)$being continuous, which in turn gives a convenient way to defining VaR and CVaR.

VaR will in this setting get the following definition which is equivalent to definition 4.3.

$$\zeta_\alpha(w) = min\{\zeta \in \mathbb{R} : \Psi(w, \zeta) \geq \alpha\} \tag{5.2}$$

which again changes the CVaR definition to

$$\phi_\alpha(w) = \frac{1}{1-\alpha} \int_{f(w,r)\geq\zeta_\alpha(w)} f(w, r)p(r)dr \tag{5.3}$$

5.1 Methods of Rockafellar and Uryasev

The main reason I present the second definition of both VaR and CVaR comes from Rockafellar and Uryasev (2000). Instead of optimizing VaR and CVaR directly, they define a similar function which is possible to linearize. They also prove that a solution to this function optimize CVaR and finds the resulting VaR simultaneously. The proof behind this function is quite extensive, and can be found in the appendix of their article.

Definition 5.1. *Given $\zeta_\alpha(w)$ and $\phi_\alpha(w)$ defined as in (5.3) and (5.5) we define the function $F_\alpha(w,\zeta)$ on $\mathcal{X} \times \mathbb{R}$ as*

$$F_\alpha(w,\zeta) = \zeta + (1-\alpha)^{-1} \int_{r\in\mathbb{R}^n} [f(w,r) - \zeta]^+ p(r)dr \tag{5.4}$$

where $[a]^+$ meaning $max(0,a)$.

When it comes to optimizing F and it's properties, Rockafellar and Uryasev prove the following theorem.

Theorem 5.1. *As a function of $\zeta, F_\alpha(w,\zeta)$ is convex and continuously differentiable. The $\alpha-CVaR$ of the loss associated with any $w \in \mathcal{X}$ can be determined by the formula*

$$\phi_\alpha(w) = \min_{\zeta\in\mathbb{R}} F_\alpha(w,\zeta) \tag{5.5}$$

In general one will always get the following

$$\zeta_\alpha \in \arg\min_{\zeta\in\mathbb{R}} F_\alpha(w,\zeta) \quad \textit{and} \quad \phi_\alpha(w) = F_\alpha(w,\zeta_\alpha(w))$$

where ζ_α is the VaR_α.

Also in a minimization problem this holds , implying we can work directly with $F(w,\zeta)$ in a risk-reward model. The functions above will be equivalent in the sense that they end at the same solution. A optimum (w^*,ζ^*) for $\min_{w\in\mathcal{X}} \phi_\alpha(w)$ will also be optimum for $\min_{(w,\zeta)\in\mathcal{X}\times\mathbb{R}} F_\alpha(w,\zeta)$. Since we have assumed continuous distribution for the losses, ζ^* will be unique and equal VaR_α, and $CVaR_\alpha$ is the minimum value. For proof of this see Krokhmal et al. (2002).

Now we have efficient way to use CVaR as a risk in optimization problem,and model (3.6) can be turned into a mean-CVaR problem.

$$
\begin{aligned}
& \underset{(w,\zeta)\in(\mathcal{X}\times\mathbb{R})}{\text{minimize}} && F_\alpha(w,\zeta) \\
& \text{subject to} && w^T e = 1 \\
& && w^T \mu \geq R^* \\
& && Lb \leq Aw \leq Ub
\end{aligned}
\tag{5.6}
$$

or alternatively

$$
\begin{aligned}
& \underset{(w,\zeta)\in(\mathcal{X},\mathbb{R})}{\text{maximize}} && w^T \mu \\
& \text{subject to} && w^T e = 1 \\
& && F_\alpha(w,\zeta) \leq R_* \\
& && Lb \leq Aw \leq Ub
\end{aligned}
\tag{5.7}
$$

Now we got an example of reward-risk model with concave (linear) reward and convex risk. This means an convex solution set and convex programming can get the optimal solution for each choice of maximum risk limit R_* and minimum return limit R^*. Varying these gives an efficient frontier as in the Markovitz case.

The underlying trouble about CVaR is that one need to compute a lot of scenarios. This means there are both a lot of constraints and variables that need to calculated and a efficient algorithm is crucial.

When applied an specific sampling, one can get an discrete version of $F(w,\zeta)$. This will be given

$$\tilde{F}_\alpha(w,\zeta) = \zeta + (1-\alpha)^{-1} \sum_{j=1}^{J} \pi_j [f(w,r_j) - \zeta]^+$$

and is proven to be convex and piecewise linear given that the continuous function is linear w.r.t w in Krokhmal et al. (2002).

The scenarios, r_j, can be sampled from either a parametric distribution with parameters calculated from earlier returns, or an empirical distribution. I will present some methods for scenario generation, but will first focus on the technical details of the model.

The non-linear term $[f(w,r_j)-\zeta]^+ = \max(f(w,r_j)-\zeta, 0)$ is now the only term left preventing this model to be LP. The solution is to replace it with auxiliary variables z_j. By defining $\tilde{F}_\alpha(w,\zeta) = \zeta + (1-\alpha)^{-1} \sum_{j=1}^{J} \pi_j z_j$ and adding the constraints $z_j \geq f(w, r_j) - \zeta$ and $z_j \geq 0$ one will get the same results as with $[f(w, r_j) - \zeta]^+$. This is all possible when scenarios are generated, and resulting in a LP problem given that the loss function $f(w, r_j)$ is linear with respect to w.

We end up with a linear representation of model (5.7)

$$
\begin{aligned}
\underset{(w,\zeta,z)\in(\mathcal{X}\times\mathbb{R}\times\mathbb{R}^J)}{\text{maximize}} \quad & w^T\mu \\
\text{subject to} \quad & w^Te = 1 \\
& \zeta + \frac{1}{1-\alpha}\sum_{j=1}^{J}\pi_j z_j \leq \omega \\
& -w^Tr + \zeta \leq z_j \qquad \text{for } j = 1, ..., J \\
& 0 \leq z_j \qquad \text{for } j = 1, ..., J \\
& Lb \leq Aw \leq Ub
\end{aligned} \tag{5.8}
$$

and similar for the risk-reward alternative. The loss function used is simply $f(w, r) = -w^Tr$ which is the negative of the amount gained from the portfolio.

The first constraint is to ensure that all money are used, so that all weights sum to 1. Note that one of these assets could be risk free.

The second constraints are the CVaR constraint. By using the J different scenarios, ζ will end up as VaR, and then take average over this quantile to calculate the CVaR. In this context where returns are used, $R^* = \omega \in (0, 1)$. Interpretation now is percent of portfolio value that is lost in the $1 - \alpha$ worst scenarios.

When optimized with prices as input, this restriction will typical be fraction of initial portfolio value. That is $\omega p^T w^0$ where p is the prices today and w^0 is initial weights. Implying CVaR will now be interpreted as monetary value of portfolio portfolio one maximum can loose in the α worst cases.

Setting $\omega = 0.15$ and $\alpha = 0.95$ means that one will maximum loose 15% of portfolio value at the average 5% worst scenarios. By increasing the fraction at risk,ω, one will get the efficient frontier as in the MV model.

The third and the fourth constraints are to ensure the model optimize over the $1 - \alpha$ worst cases, and are technical described above.

5.2 Possible constraints

For now $Lb \leq Aw \leq Ub$ has just been k possible constraints. What kind of constraints that is applied variates greatly with the portfolio manager and to whom the portfolio belong. For instance, private managers would be open to leveraged position[1], while this would be illegal for many pension funds because of the higher risk. A pension fund would try to follow a benchmark index closely, while a private manager would try to beat the benchmark with more extreme positions. All these differences can be captured through these constraints.

Position constraints
These are different constraints that affect position/weights directly. For instance, one could say $w \geq 0$ to remove short-positions. Making borrowing money (investing in risk free asset with negative value) impossible.

Second option is to say that one stock could not account for more than 20% of the portfolio value, by setting $w_i \leq 0.2$ for all assets i. This is one possible way to ensure a diversified portfolio.

Benchmark/ information ratio constraint
Information ratio is telling how much a portfolio differs from a chosen benchmark. This could for instance be a portfolio trying to beat the S&P 500 index, but with much of the same volatility. Mathematically information ratio is defined as $\frac{r_P - r_B}{Var(r_P - r_B)}$[2]. In this setting r_P is the portfolio return, and the benchmark return r_B.

The information ratio gets low if variance between portfolio returns and benchmark returns are high. Which often implies more risk since one goes against the market in order to get higher returns. If one get significant higher returns with low volatility, the information ratio is high.

This measure is used to compare performance between portfolio managers, and is for some funds limited to ensure a portfolio follows the market in terms of risk.

Transaction cost constraint
One issue with active portfolio management[3] is transaction cost. One could

[1] Leveraged position is synonymous to short position. Means borrowing money to invest more heavily in stocks

[2] More on information ratio in Kidd (2011)

[3] Active portfolio management means frequent trading with every change. Passive portfolio management means one choose some promising assets, which is kept for a longer period.

easily set up an algorithm that solved for a ”optimal portfolio" every minute, but in real life the stock exchange takes some basis points[4] for each trade.

This can be solved either by checking if expected returns increases above total cost when the position is changed or set a maximum amount of costs for on reallocation.

Given an initial portfolio w^0 with prices$p$, there is an cost per trade denoted c, we can implement a maximum total cost constraint on the form by

$$cp^T \sum_{i=1}^{N} |w_i - w_i^0| \leq \kappa p^T w^0$$

That is the differences in value of initial position and new position have to be less than some fraction κ of the portfolio value, or just number a M.

Another option is the balancing cost constraint as Krokhmal et al. (2002) uses with CVaR optimization. That is the total value of a initial portfolio have to be equal the total value of a optimized portfolio minus the cost. Technical

$$p^T w^0 = \sum_{i=1}^{N} c_i p_i |w_i - w_i^0| + p^T w$$

Both these constraints are non-linear, but can be linearized by taking $\sum_{i=1}^{N} |w_i - w_i^0|$ and change to $\sum_{i=1}^{N}(u_i^+ + u_i^-)$ with $w_i - w_i^0 = u_i^+ + u_i^-$ and where $u_i^+, u_i^- \geq 0$ for all i.

In practice fonds tend to have for instance weekly/monthly reallocation combined with constraints to minimize cost.

[4] 1 basis point = 0.01%

5.3 Scenario generation

As mentioned in the comparison between VaR and CVaR, how accurate the tail model is are of great importance for the accuracy of CVaR. The scenarios r_j with associated scenario probability π_j is therefore important. There are many methods for scenario generation, and I will present some of the conclusions from Guastaroba et al. (2009).

Mainly one can use two type of methods: parametric and non-parametric. Parametric is more used when there are less data points, and makes it possible to generate as many scenarios as needed. The downside of parametric scenario generation is that one need to assume a distribution and fit this model to data. Multivariate normal distribution is popular, but as discussed in the background theory, this go against much empirical evidence about returns. There is in general no "correct" distribution for returns.

Non-parametric generation methods don't have this problem. They have in common that there is no need to assume a distribution, but reuses historical returns. The assumptions with non-parametric is that historical values is relevant in describing future returns. Again a strong assumptions that goes again empirical research of return series.

Historical data is the easiest scenario generation method and is much used. Each data point gathered for a stock is a possible realization with equal possibility, that is $\pi_j = \frac{1}{m} \quad \forall j$ where m is number of data points. Possible alternatives are a simple rolling window average over all data to remove extremities or using for instance two weeks price gap for return(" 10 day return"), ie $r_j = \frac{P^{t+10}}{P^t} - 1$ to get a two week investment problem.

Bootstrapping is a simple resample methods similar to historical data. The method randomly picks out numbers of the data set as historical, but then replaces the number. Meaning all numbers are equally probable with same probability, but one scenario can happen more than one time. The result is that number of samples, B, can be larger than number of data points.

This is useful when there are few data points, and in non-parametric bootstrapping none assumption about distribution is made. The problem with bootstrapping in finance is that one assume the data to be independent and identical distributed, in other words uncorrelated. As we have seen, this is not the case in practice.

Block bootstrapping is a possible solution when data are correlated. Instead of drawing samples from the entire set, the data set is split into b different "blocks" of data. Each containing $k = \frac{m}{b}$ different numbers. The idea is that each block should be uncorrelated to each other, but not necessarily within each block.

This will make the resampled data to keep much of the correlation and structure within the time series of one stock, which is not the case when one pick numbers from the whole set randomly. Block size b have to be balance between correlation across different assets and correlation within the single asset. The optimal size isn't unique, but one have certain methods to find b which is close to optimal.

Monte Carolo simulation is the parametric alternative. By assuming the returns follow some distribution fitted to the data points one can easily draw as many numbers as necessary and repeat this to get a confidence interval of random numbers. Multivariate normal and multivariate student's t distribution are often used.

When considering numerical stability and sample size, some continuous parametric distributions can give an explicit solution for VaR and CVaR. For multivariate normal this using the fact that $r_P \sim \mathcal{N}(w^T\mu, \sigma_P)$ where $\sigma_P = w^T \Sigma w$, one can get a standardized normal random variable $\frac{r_P - w^T\mu}{\sigma_P} \sim \mathcal{N}(0,1)$. Now we get that

$$VaR_\alpha = q_\alpha \sigma_P - w^T\mu$$
$$CVaR_\alpha = \frac{\sigma_P}{\alpha\sqrt{2\pi}} exp\left(-\frac{(VaR_\alpha)^2}{2}\right) - w^T\mu$$

One can find similar expression for multivariate t distribution. For more see (Ch. 6 p.213 in Rachev et al., 2008)

The conclusion in Guastaroba et al. (2009) is that block bootstrap on average has the best performance. This is tested through four different periods in the stock market. "Up-up","up-down","down-up" and "down-down" where up and down means stock market as whole (big index) goes up and down.

Interestingly, student's t outperformed normal distribution in down periods, where market was heavy tailed, but was usually outperformed by normal. Indicating normal distribution may not be a bad assumptions after all.

On the other hand, when market is normally distributed the resulting optimal portfolios are equal to the mean-variance optimal portfolios. This can easily be seen in the the explicit solution above. Given that minimum return limit is binding, that is $w^T\mu = R_*$ then $\sigma_P = w^T\Sigma w$ is the only term with w. So minimizing CVaR or VaR is then the same as minimizing portfolio variance.

CVaR was proposed especially to capture downside risk more correctly than variance, and assuming normal distribution, student's t or any other symmetric distributions is somewhat to go against this purpose.

5.4 Performance measures

In the mean-variance model the natural reward-risk ratio was called Sharpe ratio, and was used to find the portfolio with most reward for each unit of risk. With both VaR and CVaR we have similar ratios, but extended to risk measures. For CVaR, and the one relevant for comparison with Sharpe in this case, is the STARR ratio(Stable Tail Adjusted Reward Ratio). Following are ideas and definitions from Stoyanov et al. (2007) and Rachev et al. (2008).

General reward risk- ratio is and the STARR ratio is defined as

Definition 5.2.

$$RR(w) = \frac{E(w^T r)}{\rho(w^T r)}$$
$$STARR(w) = \frac{w^T \mu}{CVaR_\alpha(w^T \mu)}$$

The issue with STARR ratio compared to Sharpe ratio is the difference between a risk- and a deviation measure. While standard deviation is strictly positive (unless RV X is a constant) CVaR can become both zero and negative. Since CVaR satisfied **R5**, stating $CVaR(w^T r) \geq -E[w^T r]$ which in turns imply $0 \leq -CVaR(w^T r) \leq E[w^T r]$, CVaR can become negative whenever portfolio returns are positive. In fact, a negative CVaR is actually very good performance, but makes the STARR ratio negative. Two solutions are possible:

I) Split between non-negative and negative values. Rank negative as better than positive.
II) Use a linearized form of STARR.

On a general form a linearized risk reward ratio is on the form

$$LRR(w, \lambda) = E[w^T r] - \lambda\rho(w^T r)$$
$$LSTARR = w^T \mu - \lambda CVaR_\alpha(w^T r)$$

λ is interpreted as a risk aversion parameter. Both issues in the regular STARR ratio is now fixed. Negative CVaR gives higher score, and higher ratio is better without the need for any categorization.

In Stoyanov et al. (2007) optimization of this ratio is implemented into the linear mean- CVaR model(5.8). For comparison with mean-variance model, I will stick with the full efficient frontier instead of the single portfolio resulting from max STARR method. For instance in a trading algorithm, the complete efficient frontier is of less interest and a max STARR portfolio will be computational much easier.

CHAPTER 6

Results from portfolio optimization

In this section I will show some results from mean-variance and mean-CVaR optimization. First I will show some differences in the efficient frontier between mean-variance and mean-CVaR portfolio, and when they give the same result. Secondly I want to look at how different scenario generation techniques will affect allocation and overall performance of the portfolios compared to each other.

Next I will compare some methods that can be used to optimize the mean-CVaR model with focus on computational time. Lastly I show the differences in computational effort when optimizing return with respect to risk or minimize risk with respect to return.

Mean - CVaR optimization

Period 2011 - 2016

I have based the following optimization results on weekly data of all stocks in the US based S&P 100 index in the period May 16. 2011 to May 16. 2016.

All optimizations are done with four different methods for CVaR: historical returns, block bootstrapping of historical returns, multivariate normal distribution and multivariate t distribution. All optimization of CVaR is done with $\alpha = 0.95$.

There are much variation in the efficient frontier for different type of scenarios, as we can see from Figure 6.1, with the biggest difference being how extreme the tail is. The right endpoint at the efficient frontier, given that short sale is restricted, will always be a portfolio consisting of one single stock. More precise, the stock with highest returns. This means CVaR is calculated by the tail for that single asset. In a portfolio with multiple assets there will be more scenarios in the tail, which makes CVaR less sensitive to the extreme random numbers.

The biggest difference at lower levels of risk is in the multivariate t distribution. It predicts more risk earlier at the frontier, but also higher returns towards the right endpoint.

Since multivariate t is a heavy-tailed normal distribution this is expected. It's symmetric with heavier tails, resulting in both higher returns and risk. Again, this is somewhat against empirical findings, which shows higher losses than returns, i.e asymmetric behavior described in section 2.3. For further research and asymmetric distribution could give interesting results.

The normal distribution actually shows higher risk than historical returns. Because of the asymmetry and the heavy tails implied by empirical findings in the market, this should be the other way around. Implying that the years 2011 to 2016 have been years with high returns and low risk.

Normal is still closer to historical compared to the bootstrapping and multivariate t at higher levels of risk.

It's important to note that this is just one round of sample generation. There are some variation within each scenario method that could change the results, which will be shown in the next section.

In Figure 6.2 I compare CVaR efficient frontiers with the efficient frontier from mean-variance(MV) optimization in a plot with standard deviation as risk. In this figure it's clear that the CVaR efficient frontier that is based on historical data is really close the MV efficient frontier, and thus almost mean-variance effective. At lower risk there is actually less difference in terms of variance than in CVaR for different types of scenarios, and at higher risk levels there is the same tendencies as in Figure 6.1.

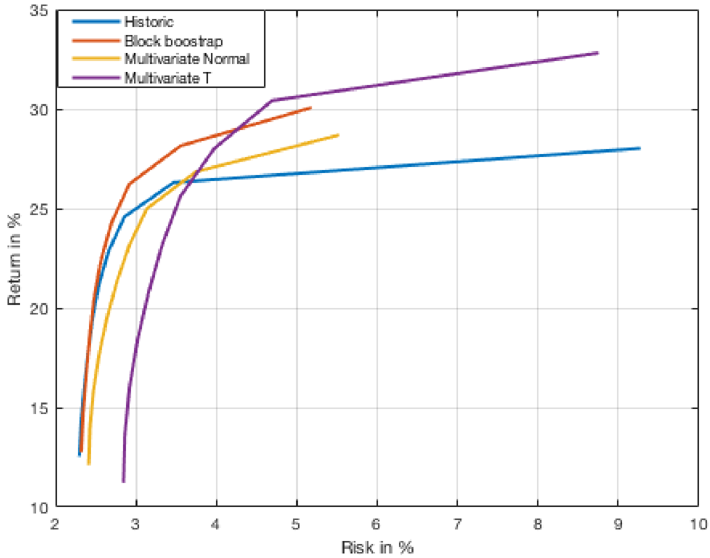


Figure 6.1: Effective frontier for CVaR optimization with different scenario generating technique in mean-CVaR space. Based on data of S&P100 stocks in the period 16. May 2011- 16. May 2016.

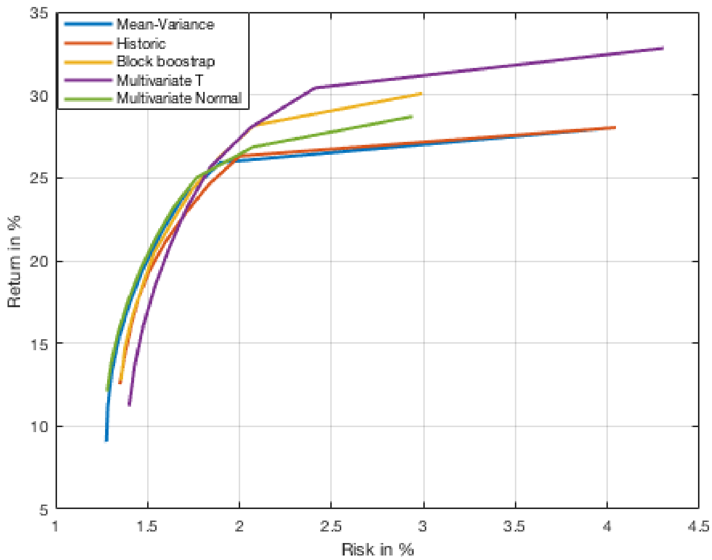


Figure 6.2: Effective frontier for CVaR optimization with different scenario generating techniques and MV optimization in mean-standard deviation space. Based on data of S&P100 stocks in the period 16. May 2011- 16. May 2016.

Period 2006 - 2011

Since 2011 to 2016 was years with less downside risk than normal, I chose to analyze the years 2006 to 2011 as well. In this period the market went into a financial crisis through 2008-2009. The S&P100 went rapidly up in 2006 and 2007, but then went down from 724.40$ to 348.21$ in the period October 1. 2007 to 1. February 1. 2009. After reaching bottom at the beginning of 2009 the index increased and by 1. of April 2011 the price was back above 2006 levels.

Because of the heavy losses and skewness in the market, CVaR should be more suited to measure risk during this period compared to variance. Somewhat surprisingly, the results of optimizing CVaR and variance are again really similar.

The effective frontier in mean-CVaR space is plotted in Figure 6.3. Multivariate t which was somewhat misplaced in good times, is now following the historical closely and predicts downsides in the market better than both bootstrapping and normal distribution. Normal distribution is to "optimistic" and predicts lower risk and higher returns that what actually is the case, as theory should suggest.

Also interesting is the fact that one could get much higher returns during the period with crisis than when the whole market went up. This probably comes from the fact that smaller stocks was more volatile and many investors searched for more secure and stable stocks.

Since the S&P 100 is composed of the 100 biggest and most established stocks in the US, these stocks would perform significantly better than smaller stocks and the market as whole. For instance the price for a stock in Amazon went from 33.78$ to 202.56$ in this period, and this is the stock with third highest return.

Looking at Figure 6.4, we see that the CVaR optimization with historical returns is surprisingly mean-variance efficient in periods of crisis as well. All frontiers are so close that randomness within the scenario generating techniques could change a conclusion. Bottom line is that optimization with CVaR is close to mean-variance efficient, but in theory more capable in describing the actual financial risk.

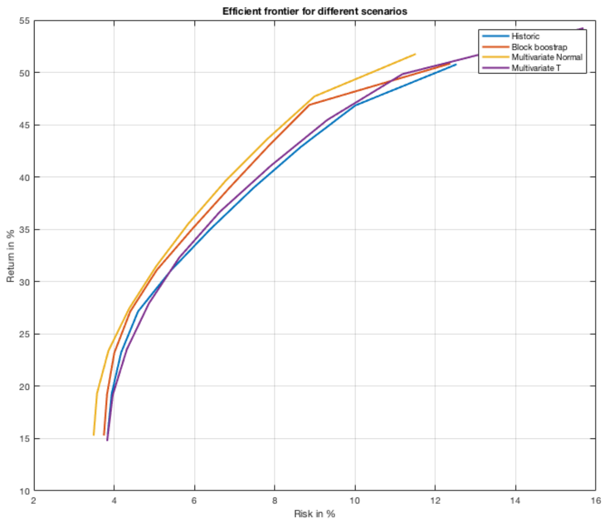


Figure 6.3: Effective frontier for CVaR optimization with different scenario generating technique in mean-CVaR space. Based on data of S&P100 stocks in the period 16. May 2006- 16. May 2011.

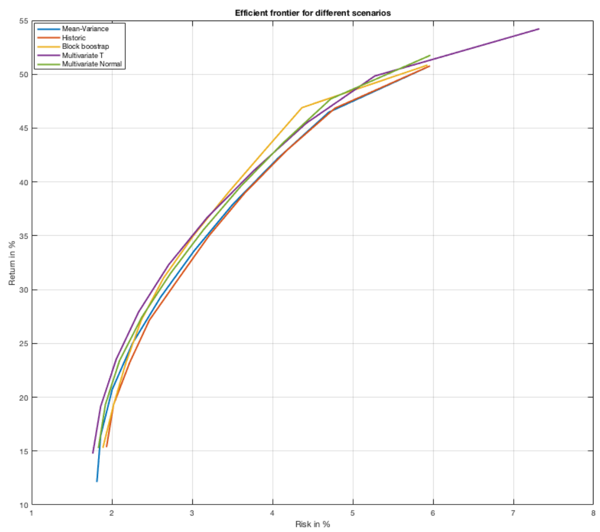


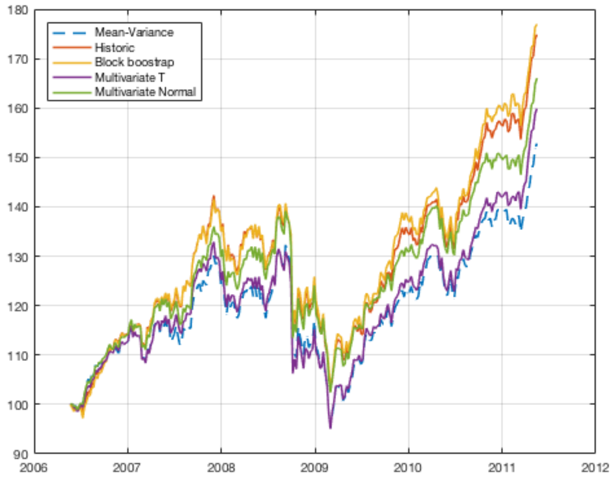
Figure 6.4: Effective frontier for CVaR optimization with different scenario generating techniques and MV optimization in mean-standard deviation space. Based on data of S&P100 stocks in the period 16. May 2006- 16. May 2011.

Mean-CVaR performance

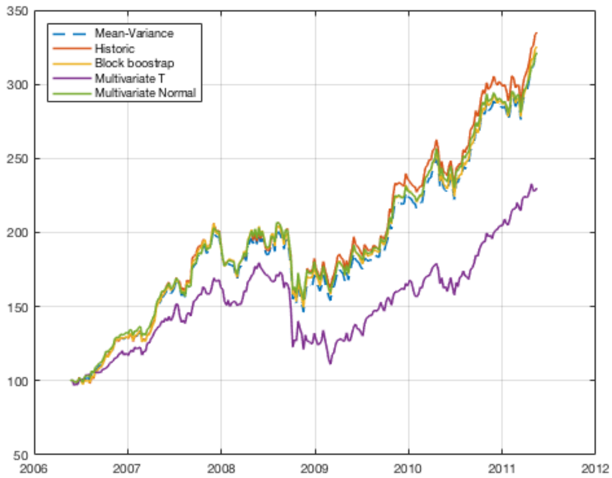
That the efficient frontiers are close implies that the weight allocation are somewhat similar. Most alike was weight allocation from mean-variance and CVaR portfolio with scenarios from multivariate normal distribution. This is what theory states and is clear from the performance results. All have similarities with which assets is chosen, but also differs quite much at lower levels of risk.

Maximum STARR ratio was found at around 5.5% CVaR during the period 2006-2011. Above this threshold there was little variation in allocation, and almost identical performance with the exception of multivariate t. Figure 6.5 shows one safe portfolio, CVaR limit at 3%, and the portfolio where most had maximum STARR ratio with CVaR around 5.5%.

The performance of a portfolio based on weight allocation from the non-parametric scenario generation techniques performed best, and multivariate t worst of the CVaR models. Mean-variance usually performed worse than CVaR optimization, but not because of less downside. It's actually when market goes up that a portfolio based on CVaR perform better. One reason for this is that variance "punishes" returns as well as losses, and it would be interesting to test with semi-variance as proposed by Markowitz.



(a) CVaR less than 3% portfolio



(b) CVaR less than 5.5% portfolio

Figure 6.5: Performance of a portfolio in the period 20. May 2006 to 20. May 2011.

Out of sample performance test

An investor will always analyze the historical data to make an investment for the future. In the following results weights are based on optimization of the data from 2006-2011, and kept in the period 2011 to 2016. By looking at the performance of keeping these stocks in this period which is outside of the historical data, one will get a sort of "out of sample" test. The results are illustrated in Figure 6.6.

Again non-parametric scenario generating techniques have the best performance, but interestingly mean-variance and normal based CVaR allocation performs better at the end of the period.

For the safe portfolio multivariate t performed best. Also in test with more strict probability level, $\alpha = 0.99$, multivariate t often did better than the others because of its heavy tails. For higher level of risk multivariate t was surprisingly off multiple times, which leads to bad performance in both in and out of sample.

The performance for multivariate t illustrated is representative for my total impression of multivariate t. It could be capable, but is often too much affected by extreme events leading to sub optimal portfolios. From my point of view, basing investment decision on a combination of normal distribution and bootstrapping or historical would be an more reliable option.

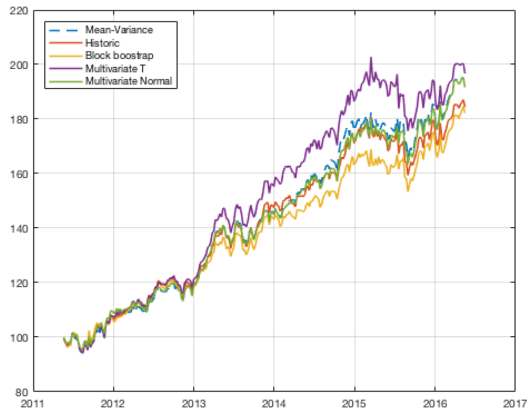
It is hard to come to an final conclusion because of the randomness connected to each scenario generation methods. All portfolios shows similarities and CVaR doesn't show any immediate advantages over variance out of sample. With the simplicity of the Markowitz model, one can understand why it's a tool many uses. Combined with robustness of mean and covariance it's still a strong competitor to newer risk-reward models.

On the other hand, measures building on the ideas of controlling extreme events are still emerging. Measures like mixed-CVaR could control absolute loss by setting α close to 1, CVaR at multiple probability levels and expected value of all losses by setting α close to 0.5 simultaneously. This in one measure that can be applied directly to the general risk-reward models presented in this thesis.

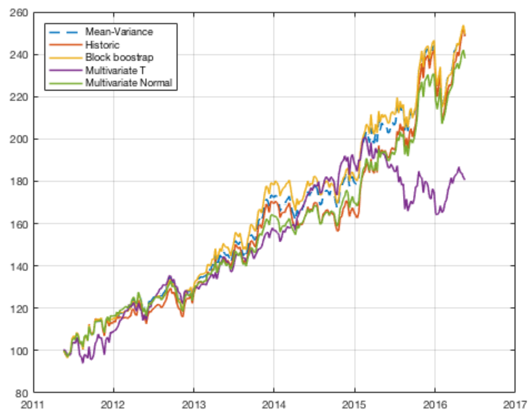
CVaR as measure is made to capture the asymmetry in the market, but in this thesis only symmetric distribution are used. Block-bootstrapping should have similar skewness as the market, but other asymmetric distribution could improve the performance of a portfolio optimized with CVaR compared to variance.

An important precision is that this is by no mean a backtest. For instance by changing allocation weekly to maximize STARR ratio through the same period would probably give an different result entirely. As stated in Rachev et al. (2008), backtesting with CVaR optimization have several difficulties.

What is done in this experiment is allocating once based on historic data and keep that allocation for five years.



(a) CVaR less than 3% portfolio



(b) CVaR less than 5.5% portfolio

Figure 6.6: Performance of a portfolio in the period 20. May 2011 to 20. May 2016.

Consistency within a distribution

When drawing scenarios from a distribution or through bootstrapping there is always some randomness, even if sample size is large. The question is if this randomness is so big that it will impact the investment advices one get from a portfolio optimization.

I have not compared the numerical precision of the CVaR calculated, which one can do by measuring the fluctuations of CVaR compared with the explicit solutions for a parametric distribution. This is already done in (p.219 Rachev et al., 2008) with sample size varying from 500 to 100 000 for $CVaR_{0.99}$. Based on this result I chose 20 000 as sample size, as a sweet spot between computational effort and precision/stability of CVaR.

In these numerical results I'm more interested in the complete efficient frontier and weight allocation at multiple levels, than precision of CVaR for the different methods.

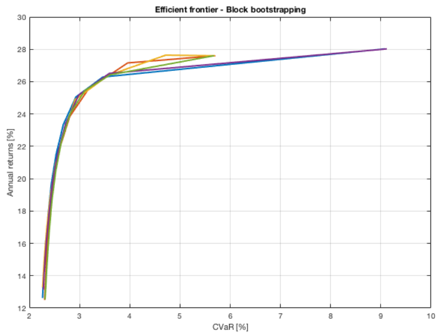
The weight allocation plots connected to each method, show three of the five optimization runs. This type of plot shows how the allocation changes from start point to end point at the efficient frontier. The "risk level" is based on the 10 return limits(R_* in model (5.6)) that is used to calculate the efficient frontier, and goes from lowest CVaR at 1 to highest possible CVaR at 10.

I have taken five runs of optimization with independent random samples from one scenario generating technique, and seen differences in the efficient frontier and weight allocation.

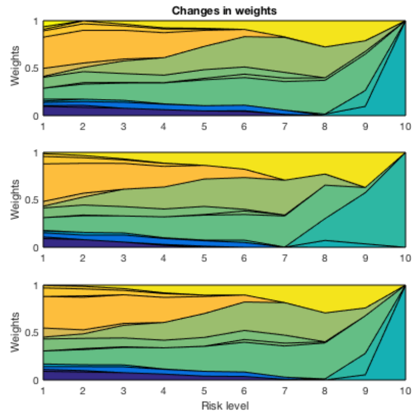
There are three assets with 27.6%, 27.9 and 28.0% expected annual return in the historical data. That the difference is so small results in all three changes between having the highest mean return after random samples are generated. This is the case for all methods.

The block bootstrapping technique is illustrated in Figure 6.7.Overall it is a stable method with resulting efficient frontiers that are close too each other. The weight allocation is consistent at the first seven risk levels, but with a tendency to be less diversified than the parametric sampling methods at higher levels of risk.

The multivariate normal samples produce more or less the same efficient frontier as can be seen in Figure 6.8. Also the weight allocation is really consistent, even at risk level 8 and almost 9.

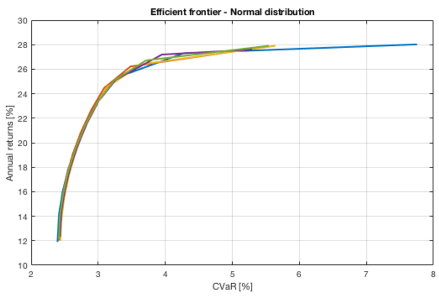


(a) The efficient frontier

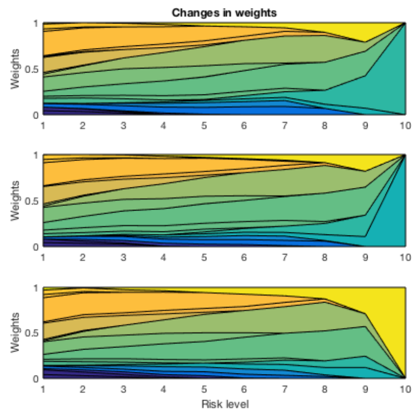


(b) Weight allocation

Figure 6.7: Five optimization runs with samples from block bootstrapping.



(a) The efficient frontier



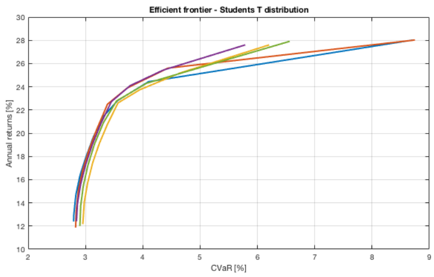
(b) Weight allocation

Figure 6.8: Five optimization runs with samples from multivariate normal distribution.

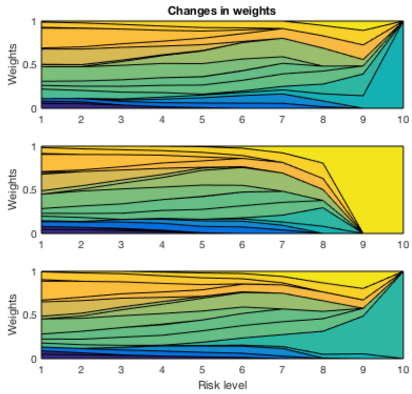
With samples from a multivariate t the variation is somewhat bigger. Looking at Figure 6.9, there are greater differences in the efficient frontier, even at the beginning, and the weight allocation changes more between risk level 7 to 9.

Notice that the allocation suggested from this optimization gets more assets on lower risk level compared to normal distribution and block-bootstrapping. This could be from the fact that all assets have heavier tails, and the bigger uncertainty results in a more diversified portfolio to lower CVaR.

From theory about diversification this should suggest lower risk and better performance, which was not the case. This could mean there is something specific with the dataset used, which makes multivariate t perform badly in this optimization, but not in general.



(a) The efficient frontier



(b) Weight allocation

Figure 6.9: Five optimization runs with samples from multivariate t distribution.

Algorithm comparison

I have used three different methods for solving the mean-CVaR optimization problem. The first method, which is the one implemented in my portfolio interface, is the prebuilt PortfolioCVaR method. By default this method uses a general solver called fmincon with a sequential quadratic programming(SQP) algorithm Mathworks calls "state of the art" in non-linear optimization.

Only requirement for this algorithm is smooth object function and constraint. Mathworks defines SQP as a medium-scale algorithm, meaning it stores all information and works on full, dense matrices. A bit more on algorithms built into Matlab are discussed in chapter 7 and full details can be found in MathWorks (2016).

The second method I used was to manually manipulating objective function and the constraint matrix into the form needed by a LP solver. This form is quite strict. Recall that a LP on standard form is defined as:

$$\begin{aligned} \text{minimize } & c^T x \\ \text{subject to } & Ax \leq b \\ & A_{eq}x = b_{eq} \\ & x \geq 0. \end{aligned}$$

where the first and the second constraint is just the coefficient matrix A split into inequality and equality constraints. For a LP solver A, b, A_{eq} and b_{eq} as well as lower and upper bounds are needed on this exact form. Looking at problem (5.8), we have to optimize the weights as well as ζ and auxiliary variables z. All these have to be one vector x with corresponding constraints added to one big matrix A.

My starting point was the objective function, i.e to maximize the expected return, $\mu^T w$. The solver takes one row given by $c^T = [\mu_1 \dots \mu_N \quad 0 \quad 0 \dots 0]$ as input. Defining the decision vector x as $x = [w \quad \zeta \quad z]^T$ resulting in x being a $(N+1+J) \times 1$ vector where N is the number of assets, and J is the number of scenarios. Now we get that $c^T x = \left[\mu^T w + 0\zeta + 0^T z\right]$

The equality constraint in this system is $\sum_{i=1}^{N} w_i = 1$. With x defined as above, one get $A_{eq} = [1 \dots 1 \quad 0 \quad 0 \dots 0]$ with $b_{eq} = 1$.

The CVaR constraint is the sum $\zeta + \frac{1}{(1-\alpha)J} \sum_{j=1}^{J} z_j$ which should be less than some limit ω. Defining $\beta = \frac{1}{(1-\alpha)J}$ one get $A_1 = [0 \dots 0 \quad 1 \quad \beta \dots \beta]$ with $b = \omega$.

The main part is the system of auxiliary variables defined by Rockafellar and Uryasev, $w^T r_j + \zeta \leq z_j$ where r_j is the scenarios generated. To get the

LP form A has to be defined $A = -\left[r_1 \dots r_N \quad \mathbf{1} \quad e_1 \dots e_J\right]$ which should be less than $b = [0 \dots 0]^T$. Here $\mathbf{1}$ is a $J \times 1$ row vector and e_i is the i'th column of the identity matrix of size $J \times J$. r_i is a row vector containing all the J scenarios for asset i.

Lower and upper bound is inserted in the solver separately. For Matlab code see section 8.2.

The last method I used was the package "Yalmip" which choose the best solver and fix constraints on proper form. This allows for constraints and variables on a more flexible form, and can call a series of different solvers outside the Matlab environment.

The state of the art solution is CPLEX developed by IBM, but because of registration and license issues I chose the competitor MOSEK optimization software. In benchmarks MOSEK is around 20% slower than CPLEX at big LP problems[1].

The problem was set with scenarios for the hundred stocks in the S&P100 index with weekly data from the period 15.may 2011 to 15.may 2016. For historic sampling this meant optimization of 262 scenarios, while the other three methods had 20 000 samples. As one can see from table 6.1, the differences in time used is quite big.

	CVaR object	Min CVaR Yalmip	Max returns Yalmip	Min CVaR LP	Max returns LP
Historic	3.5318	1.1713	1.6466	0.0527	0.1319
Bootstrapping	27.8168	4.0663	9.4258	3.9046	7.1356
Normal	24.9924	6.2656	9.2793	3.5759	7.5853
Student's t	31.0501	7.3863	10.0983	4.1916	8.6097

Table 6.1: Mean time of ten optimization runs with drawing random numbers

The built in MatLab portfolio routine is quite fast with few stocks, and have been close when I have been analyzing the German DAX index with 30 stocks. But when problem complexity is increased, LP solvers are much faster than the general optimization routines when applied to a LP problem.

As shown above, it's around 20 seconds slower than LP problems. Showing that the theory of Rockafellar and Uryasev makes an important difference. The PortfolioCVaR object works well for smaller illustrative problems, and

[1]See http://plato.asu.edu/ftp/lpsimp.html for updated benchmark of different solvers

makes more defined error messages compared to my own method. This is the reason I have used this in my interface.

For an investor that wants analysis of many stocks, together with other more complicated financial instruments, a special implemented system has to be made. This is where linearization is of great importance.

Between the two methods used for LP, Yalmip was for a long time the fastest method ans has many positive aspects. The code when programming with Yalmip get much more readable and ideas for constraints are straight forward to implement without any manipulation, see section 8.2. It's also more general, meaning it's quick to change solver if one for instance wants to add non-linear constraints.

The negative side with Yalmip is that the procedure is more hidden. When I started with the pure LP method with standard routines the total time was above 3-4 minutes. Trough optimization of the code it's now the fastest. Similar optimization could possibly be done to the Yalmip method to make a specialized version, but is hard when there are so much that is happening internally in Yalmip.

For the LP method I chose to use MOSEK's linprog[2] solver that replaces Mathworks' version. When optimizing with the regular linprog it took between 122 and 178 seconds to solve the problem compared to never above 15 seconds with MOSEK . Which illustrates well why firms like MOSEK and IBM can sell pure optimization software.

By looking at the time differences between max return versus min CVaR it's clear that even if they give the same efficient frontier, they are different in computational effort. With historical scenarios the difference are quite small. This is not surprising since the number of scenarios and stocks are quite similar, but when there are 20 000 scenarios the difference is considerable.

This difference is not one special case for one run. It has occurred when I have changed index, number of stocks, frequency and scenarios. The first thought that came to mind was difference in decision variables or rows in coefficient matrix A. Rule of thumb for dual-simplex iteration is m to $3m$ where m is number of rows in A^3. But both problems are of the same size. Matrix A in this numerical illustration has 20 001 rows and 20 099 columns, and with only the first constraint (A1 in the code) different.

[2]Name of LP solver that comes in the Matlab optimization toolbox
[3]Rule of thumb from Leon (2002)

Further I checked number of iterations which actually was less for maximizing return (17 iterations) than minimizing CVaR (22 iterations). I can just conclude that there is something within the computational work for each iterations or in how long the solvers use in phase 1 of simplex where one finds initial feasible solution. Further work is outside of this thesis.

CHAPTER 7
Portfolio optimization interface

My goal have been to implement all theoretical aspects of this thesis, in a way that should be comprehensible for someone with background in economics. I learned object oriented programming by reading Mathworks book "Creating graphical user interfaces(GUI)" and the webinars that are on their website. All code is available on `https://github.com/johalnes`.

When making a GUI the way of thinking about programming is somewhat different than mathematical scripting. Functions are often linked to certain graphical objects, with rules of name and when they are called. See chapter 8.1 for more directly connected to GUI programming.

To illustrate the extent of a GUI, my little program counts 125 objects of twelve different types(button,editable text, tables etc). Each with multiple functions and properties, individual to each type of object. Total programming will count between 2-3 000 line of codes for three separate user interfaces that are connected in one main interface.

My workflow have been something like this:

The first step was to make a sketch of classes and functions needed. I ended up with two main files. One for the GUI itself and one portfolio object, PortObj.m, which take prices etc as input and does all the financial calculations within the class, as well as initialize the portfolio functions preprogrammed in Matlab.

The second step was to make the graphical interface with buttons and other objects. For this I have used inspiration and some code from the mean-variance portfoliotool made on a webinar at Mathworks.com.

The next step was writing the code for each object in the interface and initialize the connected calculations in PortObj. This was the absolute hardest part since not all ideas made in the sketch made sense when programmed, and continuous changes had to be made.

The final step was to get all plots and tables to show the results of all calculations that was made in the third step.

The second to the final step was divided for each page. I made the import page, checked that everything worked like I wanted, then moved to the portfolio setting page and so forth.

Page 1 - Data import

The first step in portfolio optimization is getting data. The GUI has possibility for downloading the latest market data from Yahoo or importing from excel spreadsheet.

The user can choose time period, data frequency and how missing data should be treated when downloading from Yahoo.

The imported prices are shown in a table and there are two ways to further inspect the data. Either by visualizing price and returns as time series or in a distribution plot where different possible distributions are plotted with a histogram of historic returns (see Figure 7.2). The distribution plot can be used to make an idea of how scenarios should be generated later on.

In the distribution plot there is also a table with numbers of how well the distribution fits the data. This is the negative log-likelihood resulting from parameter estimation, together with Akaike information criterion (AIC) and the Bayesian information criterion(BIC).

Information criterion is used in model selection and is defined by two terms: the log-likelihood and a penalty term. The idea is that a lot of parameters may fit the data more closely, but too many gives over-parametrization[1]. With a penalty for each parameter added an information criterion will find a model with a good combination of fit and number of parameters.

The difference between AIC and BIC is that the later has more heavy penalty which usually result in less parameters. For the distributions added in this plot, over-parameterization is not an issue since all has 3 or less, and the likelihood will be the dominating term. In for instance ARMA modeling this difference is more relevant.

Note that none of these numbers can't say anything about goodness of

[1]Over-parameterizations to specified to a dataset, which leads to bad out of sample results and possible unnecessary calculations.

fit in total. They are just numbers that give sense relative too each other for a specific data set, and each data sets gives numbers of different magnitude.

Missing data occur when a stock is added or removed from a index or other reasons for a stock being temporarily stopped for sale, and happens quite often. One possibility is to remove the date from all stocks, but this could remove relevant information. By using linear or cubic extrapolation or spline for the missing data one can get more data points for use in the optimization problem.

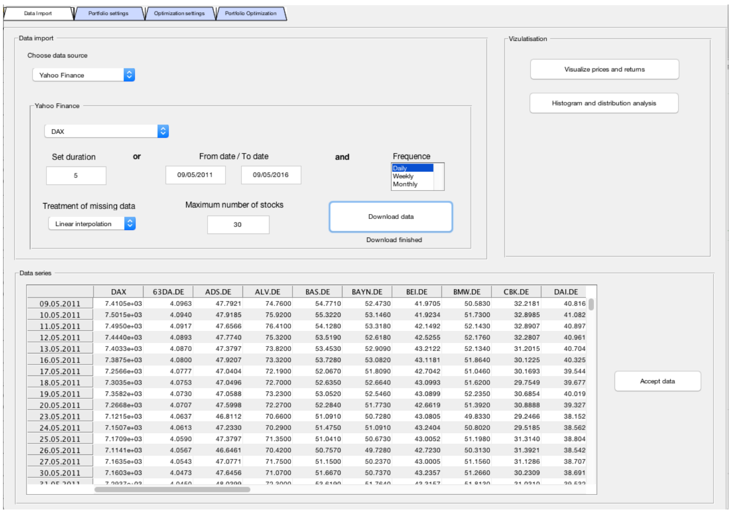
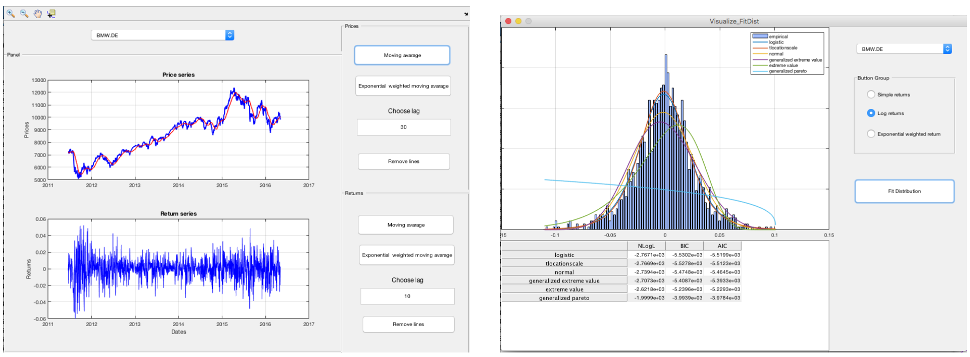


Figure 7.1: Page 1 of the portfolio interface.



(a) Price and return development

(b) Distribution plot

Figure 7.2: Illustrated with data for BMW in the period 9. May 2011 to 9. May 2016.

Page 2 - Portfolio settings

At page two of the interface are settings connected to the portfolio. The first step is to choose what kind of returns that should be used. Either simple, logarithmic or exponential weighted returns. Last is to emphasize recent returns more than old. I.e, $\sum_{i=1}^{n} \lambda^i R_{T-i}$ for $\lambda \in (0,1)$ and is equal logarithmic return when $\lambda = 1$.

The second step is to choose risk measure. Either variance, CVaR for a given α or both compared. CVaR is minimized for the probability level set, but one can examine the CVaR any probability level, α, in the result page to get a better overview of the tail risk.

The next step is to decide what type of scenarios that should be used. All scenario methods discussed in chapter 5.3 except bootstrapping are possible choices, all but historical data with 20 000 samples.

I chose to use block bootstrap with a method that finds optimal block size. The reason for this choice is that block size 1 is chosen if the returns are independent and identical distributed, but since empirical evidence shows otherwise single bootstrapping could be flawed.

The estimation of block size is according to Politis and White (2004) and coded by Andrew Patton. Politis and White propose an algorithm based on analyzing the significance of the autocorrelation[2].

By assuming a sample mean $\bar{X}$ will become asymptotically normal (central limit theorem) the long run variance can be estimated $\sigma_N^2 = Var(\sqrt{N}\bar{X}_N)$ and they show this variance as an expression based on the autocorrelation. They find an analytical expression of block size b that minimizes the mean square error of the long run variance, σ_∞^2.

The last step is to decide portfolio constraints. "Default constraint" is the constraints defined in model 5.8. That is weights sum to one (all available capital are used) and that the weights have to be positive.

Further constraints can be specified by activating a separate panel. When tracking error or turnover constraints is used an initial portfolio have to be chosen. Available quick options are equally weighted portfolio and a risk parity portfolio. In an equally weighted portfolio all assets are weighted 1

[2] Autocorrelation is describing how correlated values within a random process is at different times, for instance a highly autocorrelated price series will imply that an increase in prices today mean the prices is likely to continue to increase the day after. Autocorrelation for a set start time, but different ending times will tell how long a change in this time will impact the process.

divided by number of stocks. This is a naive implementation corresponding to the idea that stock market is completely random and analysis can't be used to predict a "optimal" portfolio. Equally weighted portfolios have actually performed quite well historical and in many cases outperforms their index.[3]

Risk parity is an idea from the 1990's that have grown in popularity after the financial crisis in 2008-2009. The idea is to focus on risk/volatility instead of returns and then use a leveraged position to get a acceptable level of returns. In my implementation the weights are the inverse of their standard deviation, $w_i = \frac{1}{\sigma_i}$, but normalized by dividing by $\sum_{i=1}^{N} \frac{1}{\sigma_i}$.

If a custom initial portfolio is wanted, the table are editable and each stock can be adjusted individually.

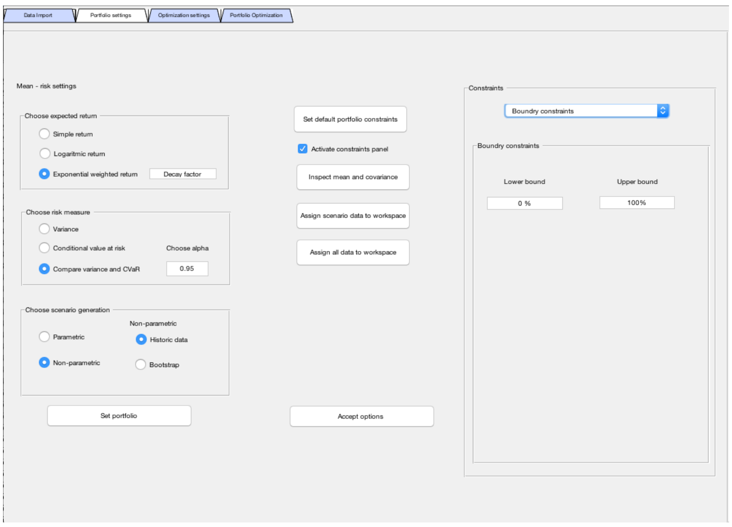


Figure 7.3: Page 2 of the portfolio interface.

Page 3 - Optimization settings

All settings connected to optimization is set at this page. MatLab have three built-in solvers for the PortfolioCVaR object: fmincon, quadprog and linprog.

Fmincon is a general convex optimization solver, and is the default solver for portfolios in MatLab. Both constraints and objective value can be non-linear and fmincon can therefore handle all kind of portfolio constraints. Fmincon can use two different types of algorithms for portfolios. Sequential quadratic programming (SQP) is the default algorithm.

[3]See thecollegeinvestor.com/8610/equal-weighted-funds-outperform-benchmark-indexes/ for an analysis on equally weighted funds that increases in popularity.

It works by iteratively looking at the Lagrangian function for the optimization problem. In each iteration the Hessian matrix[4], H, for the Lagrangian is calculated and as a start point the gradient of the objective value is used. One get a quadratic program which finds optimal direction,d, subject to the Lagrange constraints. By updating the Hessian matrix and optimizing direction one get a sequence of quadratic problems on the form $d^T Hd$ that converges to optimal solution. For further details see MathWorks (2016).

SQP is what MathWorks calls a medium-scale method. This means it stores all previous solutions, which can drain memory in large problems.

The Interior-point algorithm is more suitable for solving large scale problems, and is more memory efficient. Interior methods follows a path within the set instead of solving boundary problems, and converges towards optimal solution. This is the default solver for fmincon outside the portfolio objects.

The reason for it not being standard in portfolio optimization is that it get slightly inaccurate results in some settings. In my program this inaccuracy was enough to impact the efficient frontier, resulting in sub-optimal portfolios. In addition it was actually slower in many cases than the medium-scale methods.

Quadprog is a quadratic program solver with it's own algorithms. It requires the problem to be a quadratic programming problem, and thus more restricted than fmincon, but solves the problem much more efficient.

Because of limitations in the different algorithms, interior point is the only algorithm in quadprog general enough for optimization of CVaR. The interior point algorithm works as with fmincon, but need constraints to be convex where fmincon excepts all constraint types.

Linprog is the linear programming solver in Matlab. This solver is known to be slower than for instance CPLEX, and the speed is one of the reasons Matlab have set fmincon as default. For LP interior point method is an option, but with the inaccuracy as for the same algorithm in fmincon.

The most famous LP algorithm is the simplex method and is an algorithm that finds solutions at the boundary points of the feasible region. When there are none direction at the boundary that gives a better solution, optimum is found.

[4]The Hessian matrix is the square matrix with second order partial derivatives of a function. Note this assumes that objective function and constraints are twice differentiable. When this is the case Dahl (2010) states that the Hessian is a positive semi-definite, which gives a convex QP as introduced in the background theory.

Both quadprog and linprog uses a cutting plane method called "Kelley's method". Cutting plane methods are often used in optimization of integer problems, but is in this settings used to get convex set on a polyhedron form. This works by making an linear relaxation, that is splitting the feasible set into smaller sets by linear inequalities. By solving each of these regions, checking optimality and then again split into smaller regions on get a sequence of optimization problems that converges to the optimal solution.

With this cutting plane method linprog actually performed better with interior points. One reason could be that interior point methods in the beginning quickly converges to optimal solution where simplex goes to a boundary solution which not necessarily is in the optimal direction. This could imply less cuts and therefore less time needed.

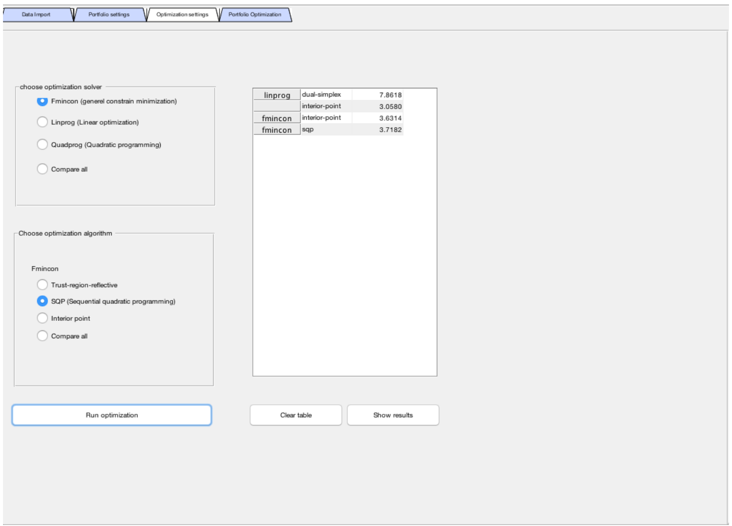


Figure 7.4: Page 3 of the portfolio interface.

Page 4 - results

This page is identical to the Portfoliotool created on one of Mathworks' webinars, but with modifications in the underlying calculations to show the theory connected to CVaR. It shows the efficient frontier first, calculated from 10 different portfolios with different level of risk and return. When a portfolio is chosen, the results for that specific portfolio are shown. This includes graphs of weights, performance of the portfolio against benchmark and VaR/CVaR, as well as the portfolio statistics.

I have calculated CVaR, STARR ratio and Sortino ratio[5] and graph of CVaR beside the metrics coded by Mathworks for the mean-variance model.

When the user chooses to compare variance and CVaR, the efficient frontier is plotted with return as y-axis and standard deviation as x-axis. This implies that the mean-variance frontier always will look most optimal when α is at 95% or higher. The user can with this setting only choose the mean-variance portfolios are the only portfolios, and therefore get statistics from from. When CVaR is chosen alone, one will have CVaR as x-axis and get statistics for the CVaR optimal portfolios.

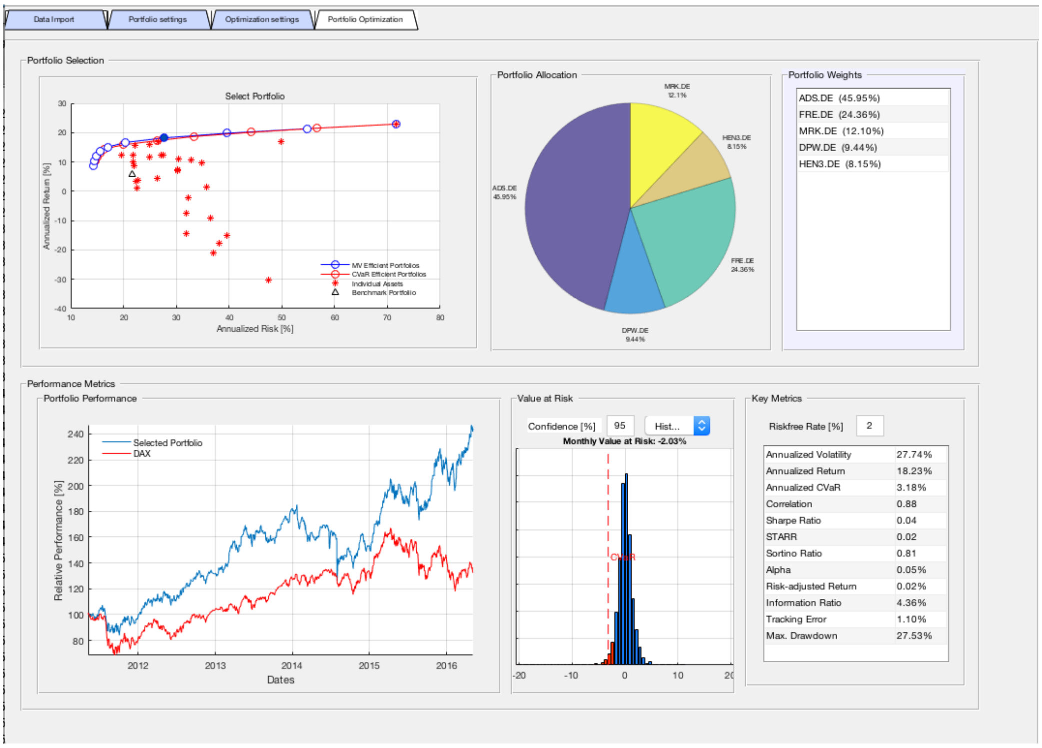


Figure 7.5: Page 4 of the portfolio interface.

[5]Same definition as Sharpe, but with lower semi-deviation instead of standard deviation as Markowitz suggested.

CHAPTER 8

Appendix

8.1 Main GUI concepts

Matlab comes with a great tool for graphical user interface(GUI) programming called GUIDE. This is a easy, "what you see, is what you get" program that makes a Matlab class with all necessary code for design. Three objects are sent between all functions in a Matlab GUI:

- *hObject :* the graphical object to the current function.
- *handles:* a collection of all the properties for all objects within the GUI. Used to change all properties like data in a table, color for all objects etc.
- *eventdata:* collection of all events that happens in the GUI.

Together these three functions holds all information and makes the user interface. The one crucial for understanding GUI programming is "handles".

The handles is the complete GUI structure/hierarchy. It stores all objects and all information. Each different type of graphical object (button, button group, editable text etc[1]) have describing properties that can be edited. Most obvious are size and color, but also options like visibility[2] or if a button is active are important. See Figure 8.1 for all properties describing a small button in a group.

[1]SeeMathWorks (2015) for all types of objects

[2]For better user experience certain objects can be invisible until for instance a choice is made.

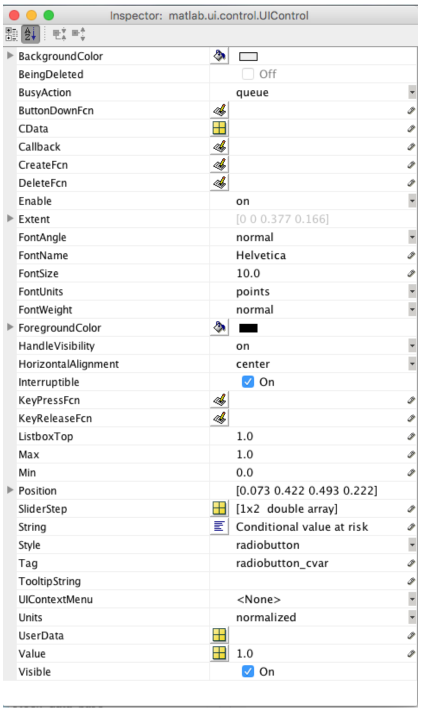


Figure 8.1: An example of all properties connected to a button.

Besides describing properties, each object has multiple functions to describe its behavior. Following are some function common for most objects.

- An *OpeningFnc* is telling what an object should to when the GUI is started. This could be that some buttons are disabled until certain actions have been made, or setting default values in text fields.
- A *CallbackFnc* is telling what an object should do when activated. For instance in a pop up-menu" which has multiple options, this is started each time the user changes option or when a button is pressed. Generally all actions connected to a object is in this function.
- A *buttonDownFcn* is started each time a mouse button is pressed when above an object. This could for instance be to reset a graph/axis when one click on it.

For a complete list of functions see page 219 in MathWorks (2015).

Following are some code from the GUI where one type duration for stocks(editable text field object):

```
function edit_duration_Callback(hObject, eventdata, handles)
```

```
% hObject    handle to edit_duration (see GCBO)
% eventdata  reserved - to be defined in a future version of MATLAB
% handles    structure with handles and user data (see GUIDATA)

duration=str2double(get(handles.edit_duration,'String'));
if isnan(duration) || isempty(duration) || (duration <= 0) || (duration > 10)
duration = 5
set(handles.edit_duration,'String','5');
set(handles.edit_stop,'String',datestr(today,24));
set(handles.edit_start,'String',datestr(today-365.25*duration,24));
else
set(handles.edit_stop,'String',datestr(today,24));
set(handles.edit_start,'String',datestr(today-365.25*duration,24));
end
```

In all functions, Matlab automatically creates descriptions of input. Notice that the eventdata is for future versions of Matlab. There have been a lot of improvements in GUIDE and other features connected to building interfaces and apps the last couple of years. With this is seems like MathWorks has more plans connected to interface programming.

The property called "tag" is an important concept. This is the name of a object in the interface which connects it with different functions. In the first line above one see this is a Callback function connected to the object "edit_duration" which is the text field's tag. I collect what the user is writing with *get(handles.edit_ duration,'String')*, but since this is in string format a conversion to number is needed to use it in calculations. Again notice that handles stores all information and to collect the information one uses *handles."tag"*.

Since this interface should be general a lot of testing is used to ensure the input is on a correct form. In the duration text box I check if the user have written a number with text(five instead of 5), only blank, negative or to big duration. If any of these test fails the text field and date fields are put back to default.

If the number is legal start and stop time are set to match the given duration. This calculated in years from the current day. If other dates than today is used this can be edited in the date fields.

The edit_duration is a really simple function, but one see that in order to build a user friendly interface all objects needs to be updated and connected. Which is all done with the commands handles, get, set and trough the tag.

8.2 Code

Mean-CVaR with Linprog

```
function [wMin,VaR,Return]=maxRetPort(S,CVaR_Lim,alpha)

[N,M]=size(S);

%Matrix manipulation to get on LP form
A1 = sparse([zeros(1,M),1 ,1/(1 - alpha)*1/N*ones(1, N)]);
A2 = -S;
A3 = -ones(N,1);
A4 = -speye(N,N);
A = sparse([A1; A2 A3 A4]);
Aeq =sparse([ones(1,M) zeros(1, N +1)]);
b = sparse([CVaR_Lim; zeros(N,1)]); beq = [1];

%Upper and lower bound
UB = sparse([ones(1,M) +Inf*ones(1,N+1)]);
LB = sparse([zeros(1,M) zeros(1, N+1)]);

%Ojbective : Maximize return for given level CVaR
objfun=-sparse([mean(S) zeros(1,N+1)]);

%Optimizing with MOSEK LP solver
options = mskoptimset('Simplex','on');
[w,fval,exitflag,out]= linprog(objfun,A,b,Aeq,beq,LB,UB,[]);

%Getting wanted results from optimization solution
wMin=w(1:M); VaR=w(M+1); Return=-fval;
```

See the use of Mosek solver and the use of Matlab's "sparse" commands[3]. When I first wrote this code, I used regular dense matrices. With 20 000 samples it took Matlab 8 to 9 seconds to just make a negative identity matrix, and it took 10 seconds to make the coefficient matrix A in the LP problem. By just adding sparse commands and use "speye" to create the identity matrix this total time went down to around 0.08 seconds.

[3]Skips zeros and only works with non-zero elements.

Mean-CVaR with Yalmip

```
function [wMax,VaR,Ret]=MaxRetPort_Y(S,Mu,CVaR_lim,alpha)

[N,M]=size(S);
w=sdpvar(M,1); z=sdpvar(N,1); R=sdpvar; zeta=sdpvar(1);

C1=[sum(w) == 1];
C2=[z >= -S*w - zeta];
C3=[zeta + (1./((1-alpha)*N)) * sum(z) <= CVaR_lim];
C=[C1, C2, C3,w>=0,z>=0];

obj=Mu' * w;

opt=sdpsettings('Solver','mosek','verbose',0);
optimize(C,obj,opt);

wMax=value(w); VaR=value(zeta); Ret=value(obj);
```

In Yalmip sdpvar stand for symbolic decision variables and is declared for optimization. As one can see, the constraints is on the exact same form as model (5.8). Yalmip is superior in readability and simplicity, with each variable optimized instead of one long matrix. Combined with for instance CPLEX it's a powerful tool for optimization in Matlab.

Bibliography

Artzner, P. and Delbaen, F. *at a1. Thinking Coherently*, volume 11. 1997.

CBinsights. The future of fintech and banking: Global fin tech investment triples in 2014, 2015. URL `https://www.cbinsights.com/blog/fintech-and-banking-accenture/`.

Dahl, G. *An introduction to convexity.* University of Oslo, 2010.

Devore, J. L. and Berk, K. N. *Modern mathematical statistics with applications.* Cengage Learning, 2007.

Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. An academic response to basel 3.5. *Risks*, 2(1):25–48, 2014.

Guastaroba, G., Mansini, R., and Speranza, M. G. On the effectiveness of scenario generation techniques in single-period portfolio optimization. *European Journal of Operational Research*, 192(2):500–511, 2009.

Kidd, D. The sharpe ratio and the information ratio. *Investment Performance Measurement Feature Articles*, 2011(1):1–4, 2011.

Krogstad, H. E. Karush-kuhn-tucker theorem. 2012. URL `https://www.math.ntnu.no/~hek/Optimering2012/kkttheoremv2012.pdf`.

Krokhmal, P., Palmquist, J., and Uryasev, S. *Portfolio optimization with conditional value-at-risk objective and constraints*, volume 4. 2002.

Krokhmal, P., Zabarankin, M., and Uryasev, S. Modeling and optimization of risk. *Surveys in Operations Research and Management Science*, 16(2): 49–66, 2011.

Landsman, Z. M. and Valdez, E. A. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4):55–71, 2003.

Leon, S. L. *Optimization theory for large systems*. Dover, New York MATH, 2002.

MathWorks. *Creating Graphical User Interfaces*. The MathWorks, Inc., 2015. URL `http://www.apmath.spbu.ru/ru/staff/smirnovmn/files/buildgui.pdf`.

MathWorks. Constrained nonlinear optimization algorithms, 2016. URL `http://uk.mathworks.com/help/optim/ug/constrained-nonlinear-optimization-algorithms.html#brnox01`. [Online; accessed 9-May-2016].

Politis, D. N. and White, H. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

Rachev, S. T., Stoyanov, S. V., and Fabozzi, F. J. *Advanced stochastic models, risk assessment, and portfolio optimization: The ideal risk, uncertainty, and performance measures*, volume 149. John Wiley & Sons, 2008.

Rockafellar, R. T. and Uryasev, S. Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42, 2000.

Rockafellar, R. T. and Uryasev, S. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*, 26(7):1443–1471, 2002.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. Generalized deviations in risk analysis. *Finance and Stochastics*, 10(1):51–74, 2006.

Sarykalin, S., Serraino, G., and Uryasev, S. Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials in Operations Research. INFORMS, Hanover, MD*, pages 270–294, 2008.

Stoyanov, S. V., Rachev, S. T., and Fabozzi, F. J. Optimal financial portfolios. *Applied Mathematical Finance*, 14(5):401–436, 2007.

Tsay, R. S. *Analysis of financial time series*, volume 543. John Wiley & Sons, 2005.

Uryasev, S., Theiler, U. A., and Serraino, G. Risk-return optimization with different risk-aggregation strategies. *The journal of risk finance*, 11(2): 129–146, 2010.

Vanderbei, R. J. *Linear programming*, volume 37. Springer, 2001.

Wikipedia. Modern portfolio theory — wikipedia, the free encyclopedia, 2015. URL https://en.wikipedia.org/w/index.php?title=Modern_portfolio_theory&oldid=688826819. [Online; accessed 16-November-2015].

Wikipedia. United states treasury security — wikipedia, the free encyclopedia, 2016. URL https://en.wikipedia.org/w/index.php?title=United_States_Treasury_security&oldid=699514365. [Online; accessed 12-February-2016].