Portfolio optimization with hedge funds

Conditional value at risk and conditional draw-down at risk for portfolio optimization with alternative investments

Author(s):
Jöhri, Stephan

Publication date:
2004

Permanent link:
https://doi.org/10.3929/ethz-a-004696440

PORTFOLIO OPTIMIZATION WITH HEDGE FUNDS: Conditional Value At Risk And Conditional Draw-Down At Risk For Portfolio Optimization With Alternative Investments

Stephan Jöhri *

Supervisor: PD Dr. Diethelm Würtz
Professor: Dr. Kai Nagel

March 16, 2004

*Master's Thesis of Stephan Jöhri written at the department of Computer Science of Swiss Federal Institute of Technology (ETH) Zurich.

Abstract

The aim of this Master's Thesis is to describe and assess different ways to optimize a portfolio. Special attention is paid to the influence of hedge funds since their returns exhibit special statistical properties.

In the first part of this thesis modern portfolio theory is considered. The Markowitz approach is described and analyzed. It assumes that the assets are identically independently distributed according to the Normal law. CAPM and APT are briefly reviewed.

In the second part we go beyond Markowitz and show that asset returns are in reality not normally distributed, but have fat tails and asymmetries. This is especially true for the returns of hedge funds. These facts justify further investigations for alternative portfolio optimization techniques. We describe and discuss therefore alternative methods that can be found in literature. They use risk measures different than the standard deviation like Value at Risk or Draw-Down and their derivations Conditional Value at Risk and Conditional Draw-Down at Risk. Based on these methods, the respective optimization problems are formulated and implemented.

In the third part we describe the numerical implementation and the used data. Finally the weight allocations and efficient frontiers that summarize the results of these optimization problems are calculated, analyzed and compared. We focus on the question how optimal portfolios with and without hedge funds are constructed according to the different optimization methods, how useful these methods are in practice and how the results differ. The results are derived by analytical work and simulations on historical and artificial data.

Acknowledgment

I would like to thank my supervisor PD Dr. Diethelm Würtz for directing this thesis and guiding me with a lot of useful impulses. I am also thankful to Prof. Kai Nagel who gave my the opportunity to work on this topic.
My gratefulness belongs also to the people at UBS Investment Research Dr. Marcos López de Prado, Dr. Achim Peijan, Laurent Favre and Dr. Klaus Kränzlein who gave me a lot of inputs during our discussions.

Contents

Part I
Modern Portfolio Theory

1 Markowitz Model

In this first chapter the fundamentals of portfolio theory are introduced. This is done by showing some statistical properties, deriving a utility function and presenting the model that combines both for portfolio optimization. This model was developed in 1952/59 by Harry Markowitz and is still considered as the standard approach for this task.

1.1 Risk Return Framework And Utility Function

Risk Return Framework

Assuming we are given N assets with their returns $R_1, ..., R_N$ respectively. Our portfolio consists of these assets with a fraction of $w_1, ..., w_N$ invested in each asset. Then the expected returns of the individual assets would be $E[R_i] = \mu_i$ (where $E[]$ indicates the expected value) and the total return μ_P of the portfolio

$$\mu_P = \sum_{i=1}^{N} w_i \mu_{R_i} \tag{1}$$

Two properties of the mean value that will become useful later:

$$\begin{aligned} \mu_{R_i+R_j} &= \mu_{R_i} + \mu_{R_j} \\ \mu_{cR_i} &= c\mu_{R_i} \end{aligned}$$

The first property means that the mean of the sum of two return series i and j are the same as the mean of return series i plus the mean of return series j. The second property states that the mean of a constant c multiplied with a return series is equal to c times the mean of the return series i.
The variance of the portfolio will be

$$\sigma_P^2 = E[(R_P - \mu_P)^2] = \sum_{i=1}^{N} w_i (R_i - \mu_P)^2 = \sum_{i=1}^{N} (w_i \sigma_i)^2 + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_i w_j \sigma_{ij} \tag{2}$$

So in the case of three assets we get the following pattern:

$$\sigma_P^2 = (w_1\sigma_1)^2 + (w_2\sigma_2)^2 + (w_3\sigma_3)^2 + 2w_1w_2\sigma_{12} + 2w_1w_3\sigma_{13} + 2w_2w_3\sigma_{23}$$

These variances $\sigma_i^2 = E[(R_P - \mu_P)^2]$ and covariances $\sigma_{ij} = E[(R_i - \mu_i)(R_j - \mu_j)] = \sigma_{ji}$ are collected in symmetric matrix called covariance matrix:

$$C = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2N} \\ \vdots & & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{pmatrix} \tag{3}$$

The correlation is defined as the standardized covariance:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \tag{4}$$

Comparing (1) and (2) we can see the **effect of diversification**: The return of a portfolio can never be smaller than the smallest return of its constituents, since it is the weighted average return of all constituents. In contrast, the variance of a portfolio can be smaller than the smallest variance of its individual assets because of the second term of (2) which can be negative in case of a negative covariance between the asset returns. So the aim of diversification is to chose the assets in a way to keep the mean return high and lower the variance by an appropriate selection and weighting of the assets.

Taking (2) with equal amount of investments in each of the N assets we get

$$\sigma_P^2 = \frac{1}{N}\overline{\sigma_i}^2 + \frac{N-1}{N}\overline{\sigma_{ij}} \tag{5}$$

whereby the first term is called **diversifiable or non market risk** and the second term **systematic or market risk**. If we take a large amount of different assets (N approaching infinity), the portfolio risk gets reduced to the average covariance of the assets in the portfolio and all the variances of the assets disappear.

$$\sigma_P^2 \xrightarrow[N \to \infty]{} \overline{\sigma_{ij}}$$

This effect shows us that the first term in (5) is called diversifiable risk because it can be reduced to zero by a good diversification of the assets. The risk represented in the first term of (5) has its origin in the risk of the single assets the portfolio contains, whereas the risk expressed in the second term is coming from the market itself (which can be influenced by economic changes or events with a large impact) and can not be reduced.
This also means that the risk of a portfolio of assets with a low correlation can be more reduced than the risk of a portfolio existing of highly correlated assets. In practice this results in the recommendation to choose the constituents of a portfolio from different geographic or industrial sectors, because assets of companies from the same country or business areas tend to move together and have hence a higher correlation. Figure 1 shows an example exhibiting this effect for the case of securities from the UK and the US.

In a risk return framework a high risk gets usually compensated by a high expected return. This is called **risk premium**: The extra return a particular asset has to provide over the rate of the market to compensate for market risk. The drawback of diversification is that the investor looses the risk premium that a certain asset might provide since its contribution on the final portfolio return is very small. The advantage of a well diversified portfolio however, is that one can expect a more moderate but constant return on the long run.

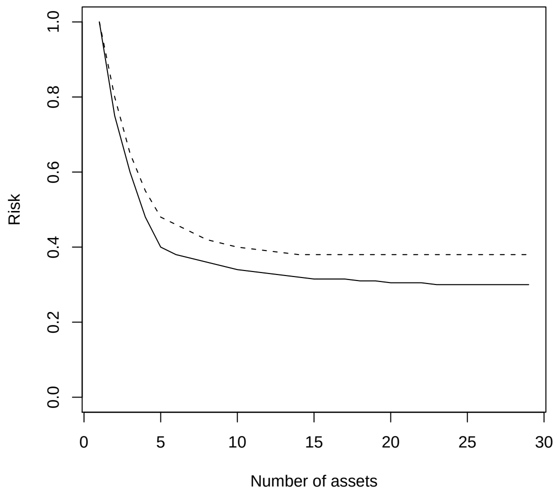


Figure 1: This chart shows the risk of a portfolio versus the number of assets the portfolio contains. We can see that a portfolio with few assets has a higher risk than a portfolio with lots of assets (effect of diversification). The doted line represents a portfolio consisting of stocks from the UK whereas the solid line depicts a portfolio with US stocks. Since the line for the UK portfolio is higher, we can conclude that the stocks in UK have a higher average covariance and their risk can therefore less reduced in a portfolio as the risk of a portfolio consisting of US stocks.

Utility Function Of An Investor

Bernoulli proposed in [9] that the value of an item should not be determined by the price somebody has to pay for it but by the utility that this item has for the owner. A classical example would be that a glass of water has a much higher utility for somebody who is lost in the dessert than for somebody in the civilization. Although the glass of water might be exactly the same and therefore its price, two persons in the mentioned situations will perceive its value differently.

We will now discuss the properties that such an utility function should have and look at some typical economic utility functions. The structure of this section will partially follow the one in Elton&Gruber [18].

The first property we want to have fulfilled is that an investor prefers more to less. Economists call this the **non-satiation attribute**. It expresses the fact that an option with a higher return has always a higher utility than an option with a lower return assuming that both options are equally likely. Or as a shorter expression, everybody prefers more wealth than less wealth. From this we can conclude that the first derivative of the utility function always has to be positive. Our first requirement for a utility function $U()$ for a wealth parameter W is therefore

$$U'(W) > 0$$

As a second attribute we want to include the investors risk profile. Bernoulli uses a fair gamble to introduce this concept. A fair gamble is a game where the expected gain is equal to zero. This means that the probability of a gain times the value of the gain is equal to the probability of a loss times the loss in absolute terms. To toss a coin would be a fair gamble if one player wins both investments when one side is up and the other payer wins both investments when the other side is up. We will examine three types of risk profiles.

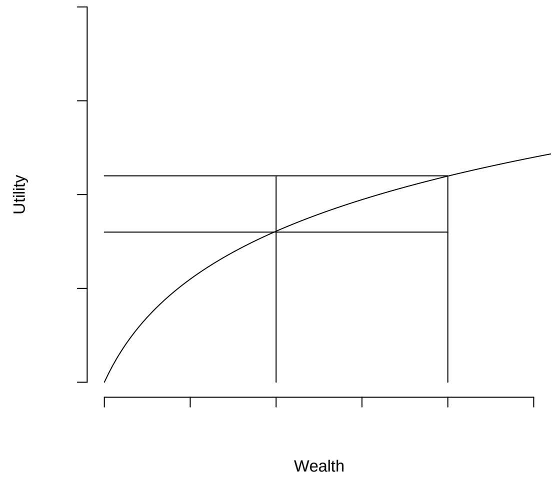


Figure 2: A logarithmic utility function seems to be appropriate in the context of a risk averse investor. The increase of the utility function for a certain increase in the wealth is smaller if the investor is already on a high level of wealth

Risk aversion is defined as rejecting a fair gamble. A risk averse investor would not play a game where he or she has an expected return of zero in the long run. Let's find out what the implications for a risk averse investor are: Since he or she does not invest, we can conclude that the utility for keeping the current wealth is higher than the probability weighted utility for a gain and loss. We can describe this risk profile for the case of a fair gamble as

$$U(W) > \frac{1}{2}U(W+G) + \frac{1}{2}U(W-G)$$

where W is the current wealth and G the symmetric gain/loss of the game.
Multiplying by 2 and rearranging yields to

$$U(W) - U(W-G) > U(W+G) - U(W)$$

and we can see that such an investor prefers the change from the current wealth minus the gain/loss to the current wealth than the change from the current wealth to the current wealth plus the gain/loss. Note that the absolute change in wealth is in both cases the same (G). From this we see that a risk averse investor prefers to keep all of his/her fortune rather than to invest a part of it and loss or gain with a 50% probability an equal part. Functions that satisfy this requirement have the second derivative smaller than zero.

$$U''(W) < 0$$

Figure 2 shows a logarithmic utility function that fulfills this property. We can see that for the double amount of wealth, the additional amount of utility is less then the double. Formulated according to the example with the fair gamble the figure expresses that for the same amount of increase in the utility the investor asks for a higher increase in the wealth the higher the wealth already is.

As second risk profile we have a look at the **risk neutral** investor. This is defined as an investor which is indifferent to a fair gamble. He or she will sometimes play and sometimes not.

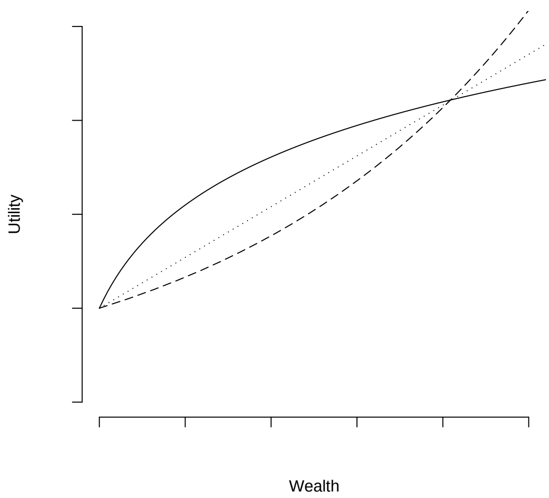


Figure 3: Utility functions for a risk averse investor (solid), risk neutral investor (doted) and a risk seeking investor (dashed) in a wealth/utility framework

For such a person the utility equation looks like

$$U(W) = \frac{1}{2}U(W+G) + \frac{1}{2}U(W-G)$$

We can rearrange this again and get

$$U(W) - U(W-G) = U(W+G) - U(W)$$

this means that such a person is indifferent about the preference of the change from the current wealth minus the gain/loss to the current wealth than the change from the current wealth to the current wealth plus the gain/loss. Hence risk neutrality causes the second derivative of the utility function to be zero.

$$U''(W) = 0$$

Risk seeking is called the third risk profile and it is defined as accepting a fair gamble. These kind of investors agree to the following formulations

$$U(W) < \frac{1}{2}U(W+G) + \frac{1}{2}U(W-G)$$

$$U(W) - U(W-G) < U(W+G) - U(W)$$

we can assign them a utility function with a positive second derivative since the wealthier they are the more they will appreciate an additional increment in their wealth.

$$U''(W) > 0$$

To conclude, in figure 3 the utility functions are drawn in a wealth/utility framework for the three risk types.

We can also transform the utility function to the Mean-Variance framework. In [27] the following utility function is proposed for this purpose

$$\mu_U = \mu_R - \lambda\sigma^2$$

where μ_U is the expected utility, μ_R the expected return, σ the standard deviation of returns and λ the risk-aversion coefficient.
With λ the function can get adapted to the investors aversion to risk. A positive coefficient indicates risk aversion, $\lambda = 0$ means risk neutrality and a negative coefficient defines a risk seeking investor. A typical level of risk aversion would be around 0.0075, as stated in [24].

It is convenient in this context to calculate the **iso-utility curves**. These curves indicate the mean/risk combinations that seem equally pleasant to a certain investor because they yield the same value for the utility function. Our three risk profile in a Mean-Variance framework are depicted in figure 4. It is possible to see how the three different types of investors get compensated: The risk averse investor (solid line) accepts a higher risk if he/she gets a higher return as compensation. The risk neutral investor (doted line) wants a certain return and does not care about the respective risk. The risk seeking investor (dashed line) accepts a return/risk combination as long as either the return or the risk is high enough. Such a person compensates high risk with low return and vise versa. From this interpretation one can see that the types of risk neutral and risk seeking investors are not very common.

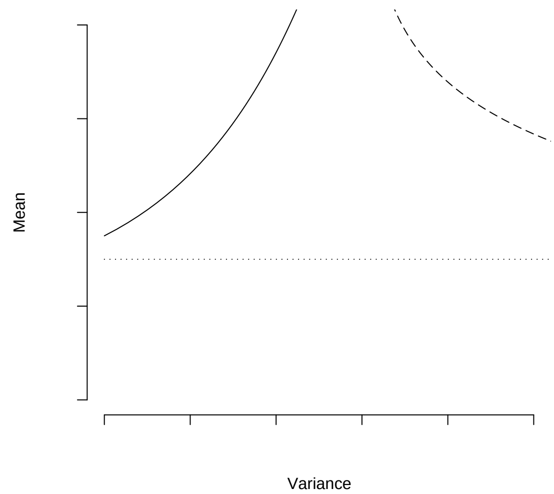


Figure 4: The Iso-Utility functions for a risk averse investor (solid), risk neutral investor (doted) and a risk seeking investor (dashed) in a mean/variance framework

There is a third property about useful utility functions that we can use to determine their appearance. It is how the size of the wealth invested in risky assets changes once the size of the wealth has changed. Again, we have three types of investors:

- Decreasing absolute risk aversion: The investor increases the amount of wealth invested in risky assets when the wealth increases.
- Constant absolute risk aversion: The investor keeps the amount of wealth invested in risky assets constant when the wealth increases.

- Increasing absolute risk aversion: The investor decreases the amount of wealth invested in risky assets when the wealth increases.

It can be shown that

$$A(W) = \frac{-U''(W)}{U'(W)}$$

measures the absolute risk aversion of an investor. As a consequence, we can define the investor types according to $A'(W)$ and assign it as follows:
$A'(W) > 0$: Increasing absolute risk aversion
$A'(W) = 0$: Constant absolute risk aversion
$A'(W) < 0$: Decreasing absolute risk aversion
It is also possible to use the change in the relative investment as property. This is expressed by

$$R(W) = \frac{-WU''(W)}{U'(W)} = WA(W)$$

and interpreted as follows:
$R'(W) > 0$: Increasing relative risk aversion
$R'(W) = 0$: Constant relative risk aversion
$R'(W) < 0$: Decreasing relative risk aversion

It is commonly accepted that most investors exhibit decreasing absolute risk aversion, but there is no agreement concerning the relative risk aversion.

In [18] two common utility functions are presented: The most frequently used utility function in economics is the quadratic one. It is preferred because the assumption of a quadratic utility function implies that the mean variance analysis is optimal (see Appendix A for a prove).

$$U(W) = aW - bW^2 \tag{6}$$

This utility function has the following first and second derivatives

$$U'(W) = a - 2bW$$

$$U''(W) = -2b$$

To make this utility function compliant to the requirements of a risk averse investor, we have to set the second derivative to be smaller than zero or b positive. We have shown that an investor usually prefers more to less and asks therefore the first derivative to be positive or $W < \frac{1}{2b}$. An analysis of the absolute and relative risk-aversion measures show that the quadratic utility function has an increasing absolute and relative risk aversion.

Since the quadratic utility function has some undesired properties, there are other utility functions in use that also satisfy mean variance analysis like

$$U(W) = \ln W$$

with its first and second derivatives

$$U'(W) = \frac{1}{W}$$

$$U''(W) = -\frac{1}{W^2}$$

It gets clear that the first derivative is positive for all values of W and the second derivative is negative for all values of W. So the logarithmic utility function also meets the requirements of a risk averse investor who prefers more to less. Further this function exhibits decreasing absolute risk aversion and constant relative risk aversion.

1.2 Selecting Optimal Portfolios: The Efficient Frontier

The basic set-up of the Markowitz [30] model is as follows:

$$w^T C w \rightarrow Min \tag{7}$$

s.t.

$$w^T \mu = \mu_P > 0 \tag{8}$$

$$w^T e = 1$$

where $e = (1, 1, ..., 1)^T$, C is the Covariance Matrix as defined in (3), μ is the expected return vector of the assets and μ_P is the desired expected return of the portfolio. The first line of the set-up defines that we want to minimize the variance and therefore the risk of the final portfolio. In the second expression we fix the expected return of the portfolio to a chosen value. It is evident that we are only interested in a return larger than zero. The last constraint sets the sum of the weights to one since we want to be fully invested.

In a short sale a trader sells an asset that is not in its possession to buy it later back and equalize its balance sheet. This practice makes sense in expectation of a decreasing price. Short sales are indicated by negative asset weights in a portfolio, since the owner of the portfolio has sold something that does not belong to him/her. If no short sales are allowed, which is usually the case, there will be an additional constraint:

$$w_i \geq 0$$

We will formulate the solution of the system according to de Giorgi [15]. Equations (7) and (8) describe a quadratic objective function with linear constraints. If the covariance matrix C is strictly positive finite, a portfolio will solve the optimization problem iff

$$w(\mu_P) = \mu_P w_0 - w_1 \tag{9}$$

where

$$\begin{aligned} w_0 &= \frac{1}{S}(QC^{-1}\mu - RC^{-1}e) \\ w_1 &= \frac{1}{S}(RC^{-1}\mu - PC^{-1}e) \end{aligned}$$

with

$$\begin{aligned} P &= \mu^T C^{-1} \mu \\ Q &= e^T C^{-1} e \\ R &= e^T C^{-1} \mu \\ S &= PQ - R^2 \end{aligned}$$

With (9) we can determine the optimal portfolio for a given expected portfolio return. This formula also sets the expected portfolio return μ_P into a relation to the portfolio variance σ_P which is

$$\frac{\sigma_P^2}{\frac{1}{Q}} - \frac{(\mu_P - \frac{R}{Q})^2}{\frac{S}{Q^2}} = 1 \tag{10}$$

A portfolio is called efficient if it offers the lowest possible risk/variance for a given expected return. The calculation of all of these optimal portfolios for different expected returns μ_P leads to set of points which called the **efficient frontier** - a hyperbola in the μ_P/σ_P^2-plane as depicted in figure 5.

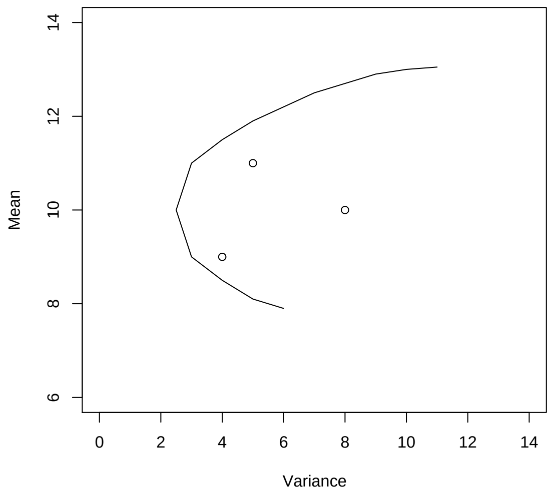


Figure 5: The efficient frontier (line) and some inefficient portfolios (points). The portfolios on the efficient frontier guarantee the highest expected return for a given variance

An important portfolio on the efficient frontier is the global **minimum risk portfolio**. It is the one to the very left of the efficient frontier. From (10) we can derive the expected return of the minimum risk portfolio as

$$\mu_{minRisk} = \frac{R}{Q}$$

From (9) we can for the global minimum risk portfolio derive

$$w_{minRisk} = \frac{1}{R} C^{-1} \mu$$

The minimum risk portfolio is the only unambiguous portfolio in the sense that there is only one possible expected return for a given variance. However, in practice, nobody will choose a portfolio lying on the efficient frontier below the minimum risk portfolio since the portfolios on the efficient frontier above the minimum risk portfolio offer a larger expected return for the same amount of risk.

With the efficient frontier we can determine the amount of risk an investor has to accept for a certain expected return he or she wants to achieve. Stated the other way around, an investor can determine, how much return he or she can expect by accepting a certain risk threshold. To define the appropriate portfolio for an investor, we can use the iso-utility curves. Figure 6 shows the efficient frontier with some iso-utility curves. The optimal portfolio is located at the point of tangency between the efficient frontier and a indifference curve (Indifference curve 2 in the example). This portfolio maximizes the utility, taking all the portfolios on the efficient frontier into consideration. Portfolios on indifference curve 3 would have a higher utility, however with the given assets we can not construct such a portfolio. Portfolios on the indifference curve 1 are achievable however not optimal in the sense of the utility.

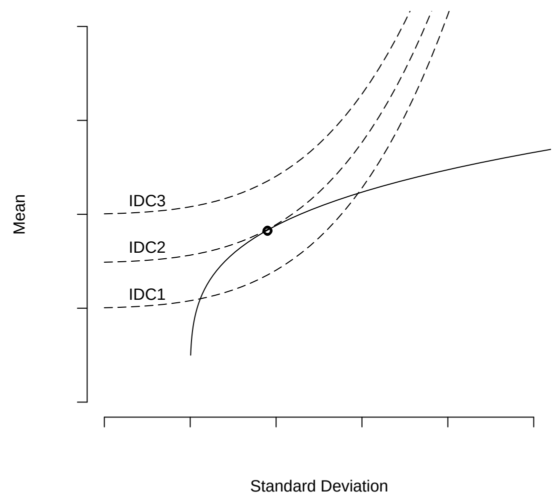


Figure 6: The efficient frontier and some indifference curves. The optimal portfolio is on the IDC2 line where the efficient frontier acts as a tangent.

In Schneeweiss shows in[39] that if one wants to apply the Mean-Variance principle as proposed by Markowitz, one has to assume that the utility function is quadratic or that the returns are normal distributed. Both requirements are critical. Not every investor needs necessarily a quadratic utility function or even a utility function in terms of mean and variance, i.e. that they chose a desired expected return and then choose the portfolio with this mean and the lowest variance. The requirement about the normality of the returns distribution will be discussed in chapter 3.

Let's follow the path of Markowitz [30] and have a closer look to the efficient frontier. In (2) we have defined the variance of a portfolio as follows:

$$\sigma_P^2 = \sum_{i=1}^{N} (w_i \sigma_i)^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}$$

since (4) holds, we can substitute σ_{ij} and get

$$\sigma_P^2 = \sum_{i=1}^{N} (w_i \sigma_i)^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \rho_{ij} \sigma_i \sigma_j$$

In the following we want to analyze the properties of the efficient frontier based on this formula for the four scenarios short sales allowed and short sales not allowed and risk-free lending and borrowing possible and not possible. For the sake of simplicity this is done for a portfolio of only two assets (i=1,2).

Short sales not allowed, no risk-free lending and borrowing

We start with the most common situation, where we are not allowed to sell assets short and no risk-free lending and borrowing is possible. Most instruments have these restrictions to avoid speculations and high risks. Three sub cases are investigated, dependent on the value of the correlation ρ between the asset returns.

Perfect positive correlation ($\rho = 1$) with $w_2 = 1 - w_1$, mean and variance of the portfolio become

$$\mu_P = w_1\mu_1 + (1 - w_1)\mu_2 \tag{11}$$

$$\sigma_P^2 = (w_1\sigma_1 + (1 - w_1)\sigma_2)^2 \tag{12}$$

It shows that with totally correlated assets, return and risk of a portfolio is just the weighted average of return and risk of its components. By solving (11) for w_1 and substituting w_1 into (12), one gets

$$\mu_P = (\mu_2 - \frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_2}\sigma_2) + (\frac{\mu_1 - \mu_2}{\sigma_1 - \sigma_2})\sigma_P$$

which is the equation of a straight line. So the efficient frontier for positive correlated assets is a linear combination of the given assets as shown in figure 7.

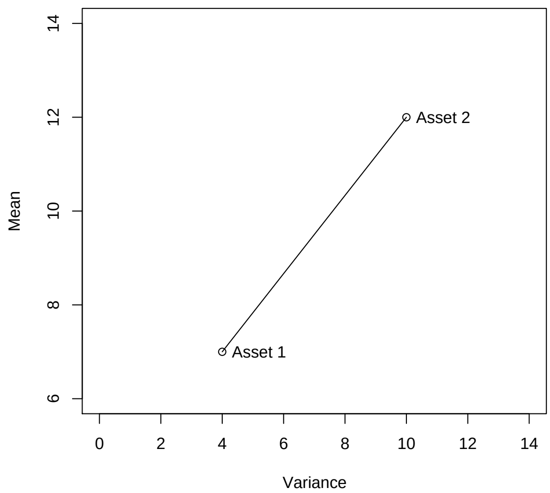


Figure 7: The efficient frontier of two assets with perfect correlation is a straight line.

Perfect negative correlation ($\rho = -1$) In the case of a perfect negative correlation, mean and variance of the portfolio become

$$\mu_P = w_1\mu_1 - (1 - w_1)\mu_2$$

$$\sigma_P^2 = (w_1\sigma_1 - (1 - w_i)\sigma_2)^2 = (-w_1\sigma_1 + (1 - w_i)\sigma_2)^2 \tag{13}$$

In the same way as in the case of positive correlation, we can find, that the efficient frontier consists of two straight lines (one for each result of the square root of (13)) drawn in figure 8. If we have perfectly anti-correlated assets, it is always possible to find combination of them which has zero risk. The appropriate weight and return can be found by setting (13) equal to zero.

$$w_1 = \frac{\sigma_2}{\sigma_1 + \sigma_2}$$

$$\mu_P* = \frac{\mu_1\sigma_2 - \mu_2\sigma_1}{\sigma_1 + \sigma_2}$$

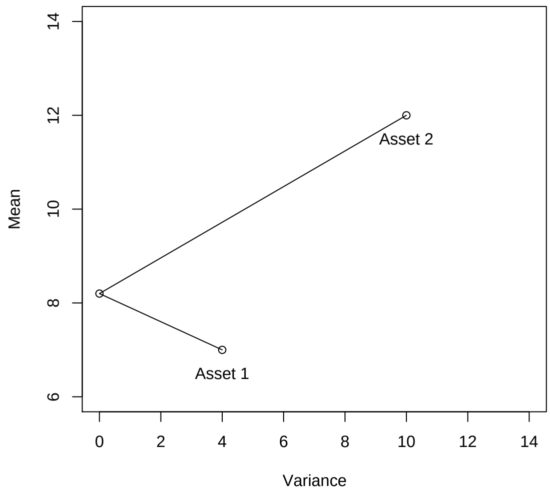


Figure 8: The efficient frontier of two assets with perfect negative correlation. It shows that the upper line has the equation $\mu_P* = a\sigma_P + \mu_P*$ whereby the lower line is $\mu_P* = -a\sigma_P + \mu_P*$ with a as a constant. The two lines intersect the y-axis at μ_P*

No relationship between returns of the assets ($\rho = 0$) For this scenario the variance of the portfolio gets simplified to

$$\sigma_P^2 = (w_1\sigma_1)^2 + ((1 - w_1)\sigma_2)^2$$

To find the minimum risk portfolio, one sets $\frac{\partial\sigma_P}{\partial w_i} = 0$ and receives for the case of two assets

$$w_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

The efficient frontier and the minimum risk portfolio are shown in figure 9.

Intermediate risk In general we can say that the efficient frontier will be always to the left of two assets, since the portfolio can be constructed as a linear combination of them. Figure 10 shows that the efficient frontier moves to the left with decreasing correlation of the assets and allows a higher diversification and therefore a lower risk.

In practice we will find almost always positive correlation between asset classes and very rarely a negative correlation. This means that there are only very few periods where a certain asset class has high profit and another asset class a negative profit. The reason lies in the factors that influence the returns of the assets classes. Most factors influence all asset classes

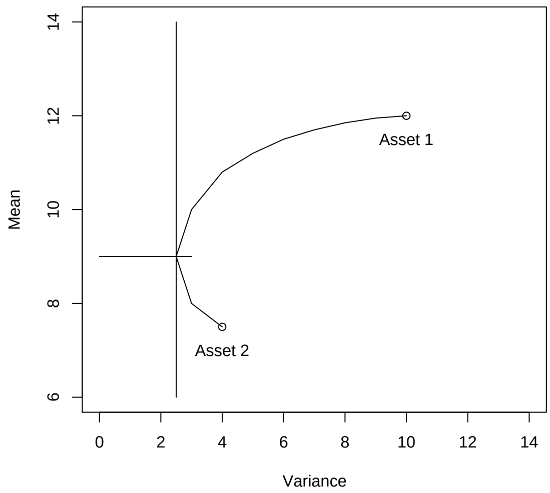


Figure 9: The efficient frontier of two assets with no correlation is a hyperbola. The minimum variance portfolio is the one portfolio at the very left

in a similar way and only a few factors influence only part of the asset classes. For this reason the behavior of the asset classes is often positively correlated.

Short sales allowed, no risk-free lending and borrowing

By doing a short sale, one takes a negative position in an asset. This may be useful in the case that one expects that the value of the asset will decrease or it might even make sense when one expects a positive return in order to get cash to invest in an asset with a better performance.

In the mean variance environment the efficient frontier will continue as a slightly concave curve to infinity. This means that one can construct a portfolio with a very high expected return by short selling a lot of assets with low expected return (see figure 11). Of course not only the expected return but also the risk of such a portfolio gets huge.

Efficient frontier with risk-free lending and borrowing

Risk-free lending is an instrument where we get a fixed interest rate μ_{rf} by lending an amount to somebody (e.g. buying government bills). Similarly, we could also get cash from somebody and pay fixed interests for it (e.g. sell government bills short). In both cases the variance of the asset is zero ($\sigma_{rf} = 0$) because the interest rates are constant. The variance of our two assets portfolio, consisting of an asset 1 and a risk-free asset rf, has a variance equal to the weighted variance of asset 1:

$$\sigma_P^2 = (w_1\sigma_1)^2$$

The optimal weight for the asset 1 would be

$$w_1 = \frac{\sigma_P}{\sigma_1}$$

As a formula for the efficient frontier we get:

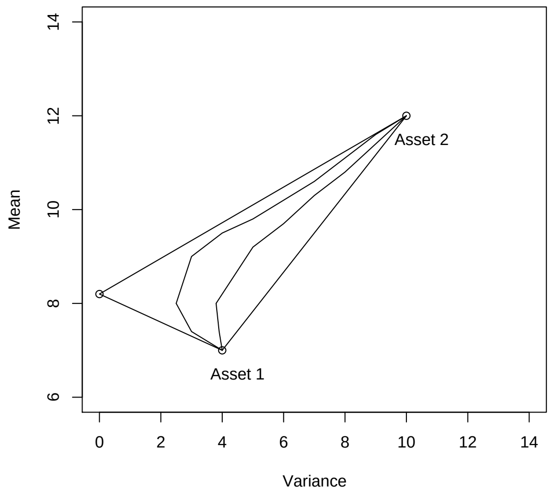


Figure 10: Comparison of the efficient frontier of assets with different correlation. The correlation of between asset 1 and asset 2 is -1, 0, 0.5, 1 (from left to right).

$$\mu_P = (1 - w_1)\mu_{rf} + w_1\mu_1 = \frac{\mu_1 - \mu_{rf}}{\sigma_1}\sigma_P + \mu_{rf}$$

From this term for the expected return of the portfolio we can see that the efficient frontier is again a linear curve as in figure 12. The term $\frac{\mu_1 - \mu_{rf}}{\sigma_1}$ or the slope of the function is called leverage factor.

To conclude, one can say that all portfolios constructed with risk-free lending and borrowing lie on one straight line through the point (μ_{rf},0) and the point representing a portfolio consisting only of the one available asset. By changing the leverage factor, one changes also μ_{rf} and σ_P^2 in a linear way.

As soon as risk-free lending and borrowing is possible, nobody will be interested anymore in the hyperbola (and its expansion through short sales) described in the section above, but only in the tangent to the hyperbola through (μ_{rf},0) since it offers a higher μ_{rf} for a given σ_{rf}.

In the case that the lending rate is not the same as the borrowing rate, we get an efficient frontier consisting out of three parts: It starts with the line of the borrowing rate until it touches the envelope of all the portfolio built without lending and borrowing and continues finally on the line of the lending rate to infinity. Since short sales allow only a concave expansion of the efficient frontier to the right and the risk-free lending efficient frontier is a straight line, short sales are also in this case of no interest anymore. An illustration is given in figure 13.

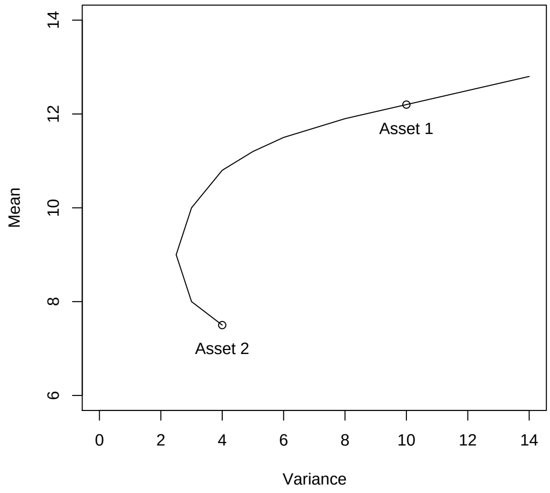


Figure 11: Short sales allow to construct portfolios with very large mean and variance because it enlarges the efficient frontier to the right.

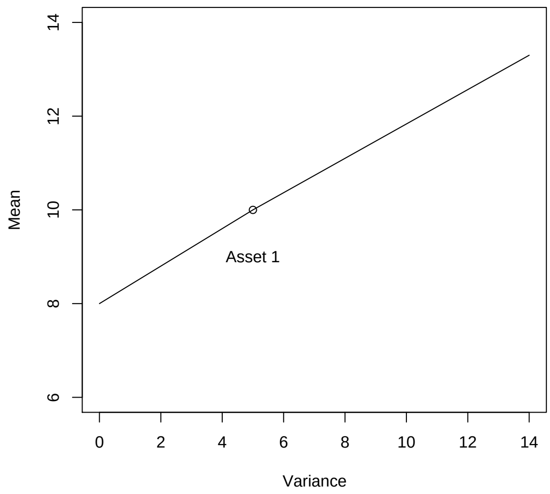


Figure 12: Risk-free lending corresponds to the efficient frontier to the left of the asset (intersection at μ_{rf} with the y-axis) and risk-free borrowing corresponds to the efficient frontier to the right of asset 1

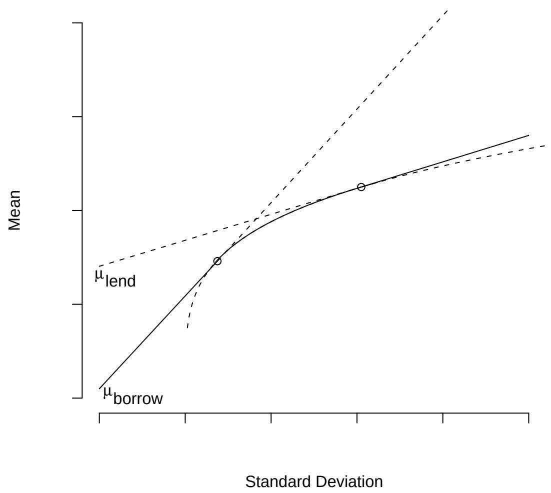


Figure 13: The efficient frontier (solid line) for different borrowing and lending rates is constructed out of three parts: First it is on the borrow line until it arrives at the hyperbola of the efficient portfolios which it follows until it reaches the tangent of the lending line where it continues to infinity.

Techniques for calculating the efficient frontier

In this chapter we will explain the techniques to determine the efficient frontier mathematically. Again, we will differentiate between the four cases of allowed and not allowed short sales and possible and not possible risk-free lending and borrowing.

Short sales allowed, risk-free lending and borrowing possible

We start with the simplest case. From the earlier chapter we already know that with allowed short sales and risk-free lending and borrowing there will be one optimal portfolio on the tangent from the risk-free asset (on the y-axis) to the envelope of all the efficient portfolios. The enabled risk-free lending and borrowing makes this tangent to the efficient frontier. Our aim is for this reason to maximize the slope of this tangent

$$\theta = \frac{\mu_1 - \mu_{rf}}{\sigma_1} \tag{14}$$

in order to maximize the return to risk ratio. There is a constraint to make sure that the weights add up to one

$$\sum_{i=1}^{N} w_i = 1 \tag{15}$$

With this setup we have a constraint maximization problem which could be solved with Lagrangian multipliers. However it is possible to turn it into an unconstraint maximization problem by combining the constraint (15) and the objective function (14). In order to do so, we start with:

$$\mu_{rf} = 1\mu_{rf} = (\sum_{i=1}^{N} w_i)\mu_{rf} = \sum_{i=1}^{N} w_i \mu_{rf}$$

Substituting this and our definition of the variance of a portfolio (2) into (14), we get

$$\theta = \frac{\sum_{i=1}^{N} w_i(\mu_i - \mu_{rf})}{\sqrt{\sum_{i=1}^{N} w_i {\sigma_i}^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij}}}$$

The maximization problem can be solved by

$$\frac{\partial \theta}{\partial w_i} = 0$$

This gives us a system of equations where we can apply the following substitution

$$z_i = \frac{\mu_P - \mu_{rf}}{\sigma_P^2} w_i$$

which leads to the following system of N simultaneous equations for N unknowns $z_1, \ldots z_N$:

$$\begin{aligned}
\mu_1 - \mu_{rf} &= z_1\sigma_1^2 + z_2\sigma_{12} + \ldots + z_N\sigma_{1N} \\
\mu_2 - \mu_{rf} &= z_1\sigma_{12} + z_2\sigma_2^2 + \ldots + z_N\sigma_{2N} \\
&\vdots \\
\mu_N - \mu_{rf} &= z_1\sigma_{1N} + z_2\sigma_{2N} + \ldots + z_N\sigma_{1N}^2
\end{aligned}$$

The optimal weights w_i can be received via

$$w_i = \frac{w_i}{\sum_{i=1}^{N} \frac{1}{2} z_i}$$

Short sales allowed, risk-free lending and borrowing not possible

If there is no risk-free asset available, we can nevertheless assume that there is a risky free asset with a specified return. Now we are in the case discussed before and can compute the optimal portfolio corresponding to this situation. By changing the return of this fictive risk-free asset to other rates, we can calculate the efficient frontier as the sum of the optimal portfolios corresponding to different rates as shown in figure 14.

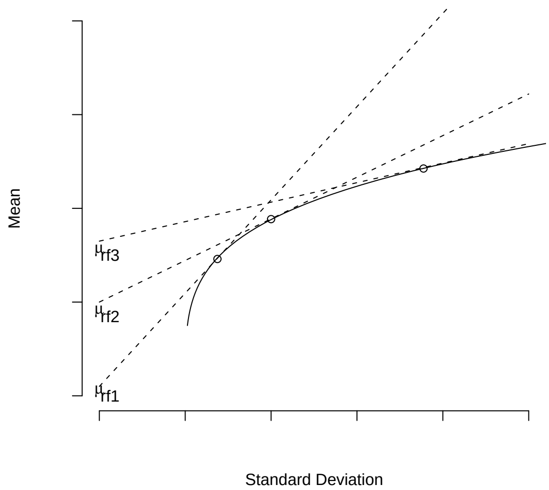


Figure 14: In the case of allowed short sales but no risk-free assets, one can determine the efficient frontier as sum of points corresponding to different (fictive) risk-free rates $\mu_{rf1}, \mu_{rf2}, \mu_{rf3}$

Short sales not allowed, risk-free lending and borrowing possible

With the restriction of no short selling, we get an additional constraint and the optimization problem looks like

$$\theta = \frac{\mu_P - \mu_{rf}}{\sigma_P} \rightarrow Max$$

subject to constraints

$$\sum_{i=1}^{N} w_i = 1$$
$$w_i \geq 0, \forall i$$

This last condition makes the problem hard to solve since we have a quadratic programming problem and no longer an analytical solution. The quadratic aspect is hidden in the objective function: The σ_P-term contains squared terms in w_i.
To solve these kind of problems, one can use a standard solver package.

Short sales not allowed, risk-free lending and borrowing not possible

If the investor does not want to allow short sales and no risk-free asset is available, we can solve the following optimization problem with the investors expected return μ_p

$$\sigma_P^2 = \sum_{i=1}^{N}(w_i\sigma_i)^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} \rightarrow Min$$

subject to

$$\sum_{i=1}^{N} w_i = 1$$

$$\sum_{i=1}^{N} w_i \mu_i = \mu_P$$

$$w_i \geq 0, \forall i$$

This is also a quadratic programming problem that should be solved with a computer package.

2 Capital Asset Pricing Model (CAPM)

This chapter presents two linear regression models to answer the question, how an efficient market behaves if every market participant follows the rules of Markowitz. The models will also be used to introduce some important concepts of finance.

2.1 Standard Capital Asset Pricing Model

The Capital Asset Pricing Model describes how a market, consisting of individual agents acting according to the model of Markowitz, behaves in the equilibrium. The Capital Asset Pricing Model has several assumptions:

- Investors make decisions solely in terms of expected value, standard deviation and the correlation structure having a one period horizon.
- No single investor can affect prices by one action - prices are determined by the actions of all investors in total.
- Investors have identical expectations and information flows perfectly.
- There are no transaction costs.
- Unlimited short sales are allowed.
- Unlimited lending and borrowing at risk-free rate is possible.
- Assets are infinitely divisible.

As we have seen above, with allowed short sales but no risk-free lending and borrowing, we get an efficient frontier like the one from A to B in figure 15. The **Separation Theorem** says that, when we introduce risk-free lending and borrowing, the optimal portfolio can be identified without regard to the risk preference of the investor (optimal Portfolio P in the figure). The investors satisfy their risk preferences by combining portfolio P with lending and borrowing and get a portfolio on the tangent to P.

According to our assumptions, all investors have homogeneous expectations and are offered the same lending and borrowing rate. In this case they will all have exactly the same diagram as figure 15. If all investors have the same diagram, they will also calculate all the same portfolio P (and variably weight it with the risk-free asset). This implies that portfolio P must be, in the equilibrium, the **market portfolio**. The market portfolio consists of all available risky assets, weighted with their market capitalization.

We can resume this and get the **Two Mutual Fund Theorem**: In the equilibrium, all investors will hold combinations of only two portfolios: the market portfolio and a risk-free security.

Figure 16 shows the market portfolio M and the same the straight line as in figure 15. This line is called **Capital Market Line**. The Capital Market Line defines the linear risk-return trade-off for all investment portfolios. It is the new efficient frontier that results from risk-free lending and borrowing. All investors will end up on it since it contains all the efficient portfolios. The equation of this line, connecting the risk-free asset and the market portfolio M, is

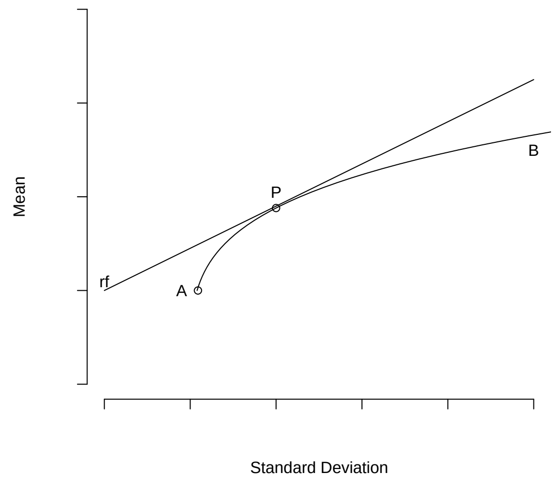


Figure 15: The efficient frontier and its tangent at the optimal portfolio. By lending and borrowing, one moves on the tangent: Portfolio P is without lending and borrowing. If one lends additional capital from somebody, one gets a portfolio on the tangent to the right of P and if one borrows capital to somebody one gets a portfolio on the tangent to the left of P.

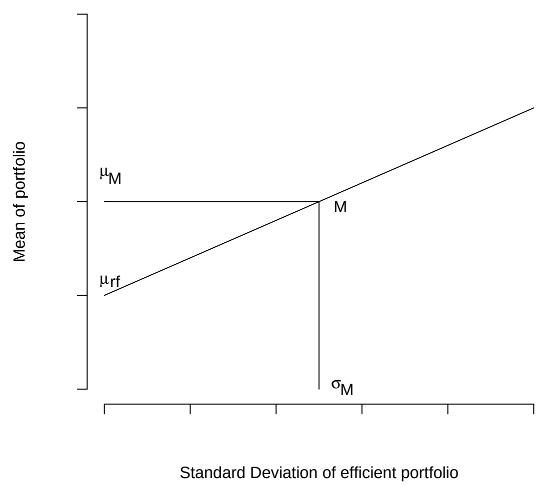


Figure 16: The Capital Market Line describes the linear relation between risk and return for a portfolio. The market portfolio is depicted as M

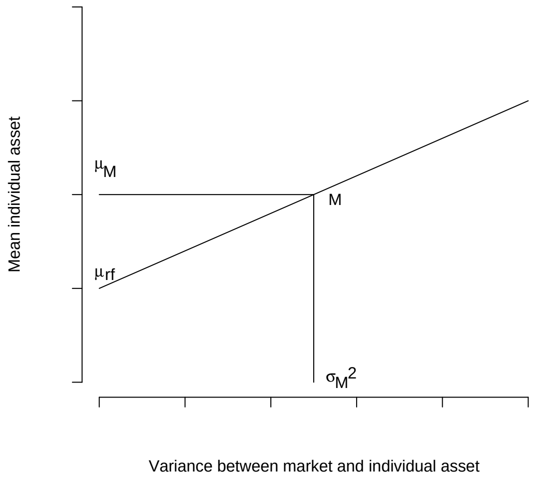


Figure 17: The Security Market Line describes the linear relation between risk and return for a portfolio.

$$\mu_P = \mu_{rf} + (\frac{\mu_M - \mu_{rf}}{\sigma_M})\sigma_P$$

This can be interpreted as

Expected return= reward for time + reward for risk * amount of risk

Let's have a look at the individual assets: The relevant measure here is their covariance with the market portfolio ($\sigma_{i,M}$). This is described by the **Security Market Line**: The Security Market Line defines the linear risk-return trade-off for individual stocks. Its formula is

$$\mu_i = \mu_{rf} + (\frac{\mu_M - \mu_{rf}}{\sigma_M})\frac{\sigma_{iM}}{\sigma_M}$$

At this point we would like to introduce a factor called **beta**. It is a constant that measures the expected change in the return of an individual security R_i given a change in the return of the market R_M. It can be estimated by

$$\beta_{iM} = \frac{\sigma_{iM}}{\sigma_M^2}$$

We can use this to substitute beta for the two variances:

$$\mu_i = \mu_{rf} + \lfloor \mu_M - \mu_{rf} \rfloor \beta_i$$

Finally we derive a single index model that describes the relation between the return on individual securities and the overall market at a time point t:

$$R_{it} = \alpha_i + \beta_i R_{Mt} + \epsilon_i \tag{16}$$

where
α_i: part of the return of security R_{it} that is independent of the market's performance R_{Mt},
β_i: sensitivity of return of security R_{it} to market's performance R_{Mt},

R_{Mt}: return of the market,
ϵ_i : a random error term with mean equal to zero.

Beta measures how sensitive a stock's return is to the return of the market. A beta of two means that the return of the stock will be the double of the return of the market (no matter whether it is a loss or a gain). Similarly, a beta of 0.5 means that the stock will move only half as much as the market does. In other words, a stock with a high beta gets a high risk premium and a stock with a low beta gets a low risk premium.
The intention of splitting the return of a stock into a part that is related to the market ($\beta_i R_{Mt}$) and a part that is related to the individual stock (α_i) comes from the observation, that when the market goes up, most stocks follow this trend and vice versa. Therefore is a part of the stock return related to the market return. It is interesting in (16) to see that the return is only influenced by the market risk and investors don't receive a premia for holding additional diversifiable/non market risk.

We can summarize that the Capital Asset Pricing Model is a theoretical model to identify the tangency portfolio. It uses some ideal assumptions about the economy to conclude that the capital weighted world wealth portfolio it the tangency portfolio and that every investor will hold this portfolio.

2.2 Arbitrage Pricing Theory (APT)

The Arbitrage Pricing theory is an alternative approach to determining asset prices. It was first introduced in [37] and bases on the idea that exactly the same instrument can not be differently priced.

As we have seen, the Capital Asset Pricing Model has some quite restrictive assumptions. This gives space for the Arbitrage Pricing Theory. It asks for the following conditions to be fulfilled

- Returns are generated according to a linear factor model.
- The number of assets N is close to infinite.
- Investors have homogenous expectations (same as in CAPM).
- Capital markets are perfect (perfect competition, no transaction costs - same as CAPM).

The Arbitrage Pricing Theory states that returns of stocks are generated by a linear model consisting of F factors I_j

$$R_i = a_i + b_{i1} I_1 + b_{i2} I_2 + \ldots + b_{iF} I_F + e_i \tag{17}$$

where
a_i : the expected return for stock i if all factors have a value of zero,
I_j : the value of factor j that impacts the return on stock i,
b_{ij} : the sensitivity of stock i's return to factor j,
e_i : a random error term with mean equal to zero and variance equal to $\sigma^2_{e_i}$. This error is uncorrelated with the factors b_{ij} and errors of the other assets (unsystematic risk).

If the assumptions hold, we can combine the assets to get a risk-free portfolio that requires zero net investment (i.e. by short selling certain assets and buying others with the revenue). The fundamental implication of the Arbitrage Pricing Theory is that such a free, risk-free portfolio (arbitrage portfolio) must have a zero return on the average. This is intuitive since a risk-free portfolio with an expected return of non zero is an arbitrage opportunity which would be exploited immediately by market participants and hence diminish.

Let's express this in a more mathematical way: Using (17) we can write the expected portfolio return as

$$\mu_P = \sum_{i=1}^{N} w_i a_i + \sum_{i=1}^{N} w_i b_{i1} I_1 + \ldots + \sum_{i=1}^{N} w_i b_{iF} I_F + \sum_{i=1}^{N} w_i e_i \tag{18}$$

We have assumed that the number of stocks are close to infinite. So, it is possible to find a portfolio that satisfies the following properties:

$$\begin{aligned}
\sum_{i=1}^{N} w_i &= 0 \\
\sum_{i=1}^{N} w_i a_i &= 0 \\
\sum_{i=1}^{N} w_i b_{i1} &= 0 \\
\sum_{i=1}^{N} w_i b_{i2} &= 0 \\
&\vdots \\
\sum_{i=1}^{N} w_i b_{iF} &= 0
\end{aligned}$$

The first condition defines that we have no net investment since we want an arbitrage portfolio. The second condition asks the expected return for this stock to be zero if all factors are set to zero (non-arbitrage condition). The following conditions imply that the portfolio has no risk since it has no exposure to any of its constituents. These three types of conditions are called orthogonality constraints. Applying them to (18), we can see that it must produce an expected return of zero. Again, if this would not hold true, investors would have a free money generator.

It shows that the orthogonality constraints imply that the expected returns μ_{R_i} are a linear combination of the b_{ij} and a constant. This means that there exists a set of factors $\lambda_0 \ldots \lambda_F$ such that

$$\mu_{R_i} = \lambda_0 + \lambda_1 b_{i1} + \ldots + \lambda_F b_{iF}$$

The b_{ij} can still be interpreted as the sensitivity of the assets to a change in an underlying factor I_i. In contrast, the λ_j represent the risk premia of the respective factor.

We are determining now the λ_j by using the fact that an asset with single exposure to one factor and no exposure to the other factors has the same risk premia as this factor. For each

$\lambda_j, j = 1 \ldots F$ we do the following: The respective b_{ij} is set to 1 and all the other equal to 0. With this procedure we find that

$$\mu_{R_i} = \lambda_0 + b_{i1}(\mu_{R_1} - \lambda_0) + \ldots + b_{iF}(\mu_{R_F} - \lambda_0)$$

We assume that for $i = 0$ we have the risk-free asset since the risk-free asset does not depend on any other factors ($b_{0j} = 0$, $j = 1 \ldots F$). For this special case of the APT model we get

$$\mu_{R_0} = \lambda_0 = \mu_{rf}$$

and therefore we can express the model as formula for the excess return

$$\mu_{R_i} - \mu_{rf} = b_{i1}(\mu_{R_1} - \mu_{rf}) + \ldots + b_{iF}(\mu_{R_F} - \mu_{rf})$$

The Capital Asset Pricing Model can be seen as a very special case of the Arbitrage Pricing Model with only one factor (single index model). This can be shown if one sets F = 1. Then we have left

$$R_i = a_i + b_{i1}I_1 + e_i$$

Now we can interpret a_i as the return of the risk-free asset μ_{rf} and $b_{i1}I_1$ as the return of the market portfolio R_M times the leverage factor.

$$R_i = \mu_{rf} + b_1 R_{Mi} + e_i$$

And this is the same expression as (16) for the CAPM.

Factor analysis is the principal methodology used to estimate the factors I_j and factor loadings b_{ij}. Since it is not possible to calculate a perfect specification of the model described by (17), a factor analysis will derive a good approximation. The criteria for the goodness is the covariance of residual returns which should be minimal. To execute a factor analysis, one has to determine the number of desired factors in advance. By repeating this process for an increasing number of factors, one gets one solution for each number of factors. A criteria to stop increasing the number of factors would be, if the probability that the next factor explains a statistically significant portion of the covariance drops below some level (e.g. 50%).

There are factor analysis methods that produce orthogonal factors (e.g. principal component analysis) and others that produce non-orthogonal factors. It may become a disadvantage to choose a method that creates orthogonal factors since the factors it creates do not exist in the real world and can therefore not be interpreted. However they can be used in a pure statistical model by assuming that the past data will be valid for the next step and applying them to calculate one step into the future. The non-orthogonal model might be not so accurate, but as soon as one gets the factors (like indices or interest rates) and their respective weights, one can apply the model in the future with new data from these factors.

To conclude, we can say that the Arbitrage Pricing Model has a number of benefits: It is not as restrictive as the Capital Asset Pricing Model in its requirement concerning the distribution of the returns and the investors utility function. It also allows multiple sources of risk to explain the stock return movements. Further it avoids using the concept of a Market portfolio. This is an advantage because this concept is hard to observe in practice.

The flexibility is also the main disadvantage of the model: The investors have to decide which sources of risk they want to include and how to weight them. Further the APT model might not be so intuitive as the CAPM.

Nevertheless, the Arbitrage Pricing Theory remains the newest and a promising explanation of relative returns.

Part II
Beyond Markowitz

We have seen in the first part that the approach to optimize a portfolio as proposed by Markowitz asks for some strong assumptions like normal distributed returns. In this second part we will investigate whether it can be assumed that the returns of financial assets are produced by a normal distribution. As we will seen, there will be several aspects that indicate that this assumption does not hold. We will use this as justification for analyzing further portfolio optimization algorithms that do not have such a strong requirement to the underlying distribution function of the asset returns.

3 Stylized Facts Of Asset Returns

In this chapter we will present some statistical tests to investigate the characteristic properties of financial market data. The used tests are chosen with respect to the properties that are important specially for financial time series. The tests for determining the form of the underlying distribution function that has created the returns are Goodness of fit (Kolmogorov-Smirnov test), Kurtosis and Skewness (Jarque-Bera test) and Quantile-Quantile plots. Concerning the form of the distribution function, we especially test for the Normal distribution. Further we have selected two tests for detecting dependencies and long memory effects in the time series. These are the Runs test for randomness and BDS test for dependencies.
The focus of the tests as a whole lies on the detection of fat tail behavior rather than dependencies. The tests are presented in their functionality and demonstrated on representative, artificial data. In part III of the thesis the tests are applied to real market data and the resulting conclusions drawn.

Non normality in return distributions

A very important question in financial analysis is the one for the distribution function of the asset returns. Since a lot of methods and theorems are assuming a certain distribution function, it is crucial to analyze the origin of the returns.
There are two aspects of the distribution function that has created the asset returns that should be considered:

- Form: Does the distribution have fat tails or skewness?
- Dependencies: Do the returns depend on an earlier return values?

The normal distribution was first mentioned by de Moivre in 1733 [31]. The advantages of this distribution are

- It can be defined by only two variables: mean and variance.
- It describes random behavior in a natural mechanisms.

For this reasons and the fact that it is possible to fit it as a first approximation to asset returns, the normal distribution is used a lot in financial analysis and is still considered as the standard assumption.
However, in 1963 Mandelbrot [29] observed that financial returns might not be produced by a normal distribution.

3.1 Distribution Form Tests

Goodness of fit test (Kolmogorov-Smirnov test)

We start with the Kolmogorov-Smirnov one-sample test which can be used to answer the question, whether a sample comes from a population with a specific distribution. The test is based on the empirical distribution function of the given samples and is restricted to continuous distributions to test for.
Assuming we are given the samples as $X_1, X_2, \ldots, X_N$. We can order them and calculate the empirical distribution function as

$$E_N = \frac{n(i)}{N}$$

with $n(i)$ as the number of samples that are smaller than X_i. The Kolmogorov-Smirnov test determines the maximum distance between this empirical distribution function and the cumulative distribution function of the assumed underlying function. Figure 18 shows a chart with these two distribution functions.

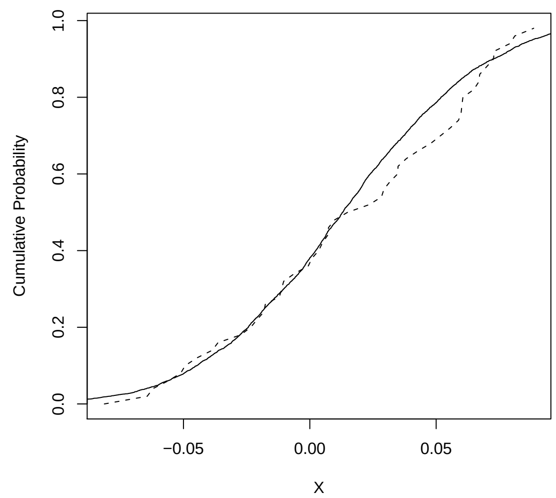


Figure 18: The Kolmogorov-Smirnov test calculates the maximum difference between the empirical distribution function of the samples (doted line) and the cumulative distribution function of the assumed underlying function (solid line).

The hypothesis of the test are defined as:
Null hypothesis: The data follows the assumed distribution

Alternative hypothesis: The data does not follow the assumed distribution
The precise test statistic is

$$D = \max_{i \leq i \leq N} |F(X_i) - \frac{i}{N}|$$

with $F(X_i)$ as the assumed underlying distribution function. The null hypothesis of the distribution is rejected if $\sqrt{N}D_N$, dependent on the confidence level, is greater than the critical value derived from the standard normal distribution.

There are two equivalent ways to handle the underlying distribution. In both ways the mean $\hat{\mu}$ and variance $\hat{\sigma}$ of the underlying distribution need to be estimated out of the given samples. It is then possible to compare the samples to a normal distribution with the estimated mean $\hat{\mu}$ and variance $\hat{\sigma}$. Otherwise one can transform the given samples according to

$$\hat{X}_i = \frac{X_i - \hat{\mu}}{\hat{\sigma}} \tag{19}$$

and compare the new samples to a standard normal distribution.

Some points classify the Kolmogorov-Smirnov test as unsatisfiable for our purpose: First, since the test compares the absolute difference between the two cumulative distributions, it underweights the difference in the tails and overweights the difference near the mean of the distribution. However we want especially check whether our distribution has fat tails. The second disadvantage of the Kolmogorov-Smirnov test is that it is a very general method (it can also be used for comparing with other distributions than just the normal) and is thus taking only the mean and variance of a distribution into consideration.

Skewness and kurtosis (Jarque-Bera test)

For the Kolmogorov-Smirnov test we were looking at the first and second moment of the distribution.

$$\mu = \sum_i w_i x_i$$
$$\sigma^2 = \sum_i w_i (x_i - \mu)^2$$

In terms of the normal distribution, often the third and fourth moments become interesting. Skewness is the standardized third moment

$$\varsigma = \frac{\sum_i w_i (x_i - \mu)^3}{\sigma^3}$$

Skewness can be interpreted as a measure for the asymmetry of a distribution function whereby a value of 0 indicates absolute symmetry (e.g. the normal distribution), a positive skewness means an increased probability at the higher quantiles (heavy right tail) and a negative skewness says that we have an increased probability at the lower quantiles (heavy left tail). Figure 19 shows some examples of empirical distributions with skewness.

The standardized fourth moment is called kurtosis. Because the normal distribution has a kurtosis of 3, one often calculates the excess kurtosis which is the kurtosis minus 3.

$$\kappa = \frac{\sum_i w_i (x_i - \mu)^4}{\sigma^4} - 3$$

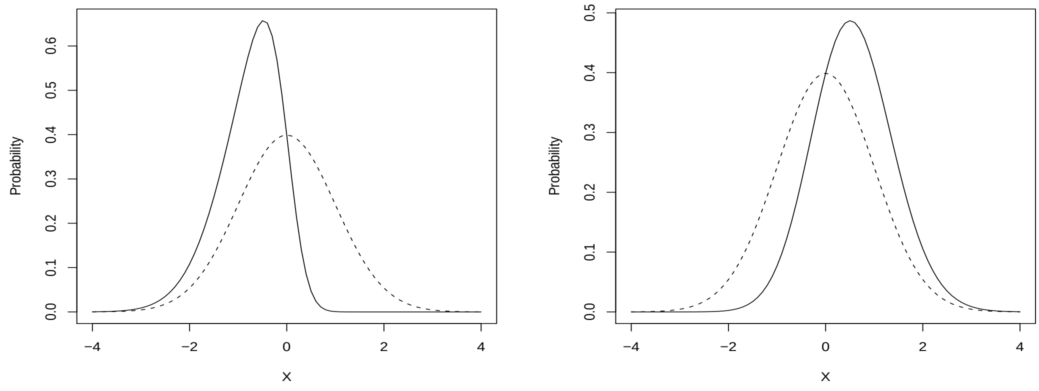


Figure 19: The charts show skewed normal distributions (solid) in comparison with a normal distribution (doted). The left chart is drawn by a standard normal distribution with a shape parameter of -3, while the right chart is drawn by a standard normal distribution with shape parameter of 1.

The kurtosis of a distribution defines whether the distribution has fat tails in comparison with a normal distribution or not. The following holds true for most financial time series: A negative kurtosis indicates that both tails are less pronounced and the distribution is less peaked as a normal distribution (platykurtic). A distribution with a kurtosis of 0 is called mesokurtic. The opposite of platykurtic, a positive kurtosis, means fat tails and more peakedness than a normal distribution (leptokurtic). If there is excess kurtosis, the mid-range values on both sides of the mean have less weight than in a normal distribution. This means that distributions with a high kurtosis are appropriate when the returns are likely to be very small or are likely to be very large but are not very likely to have values between these two extremes.

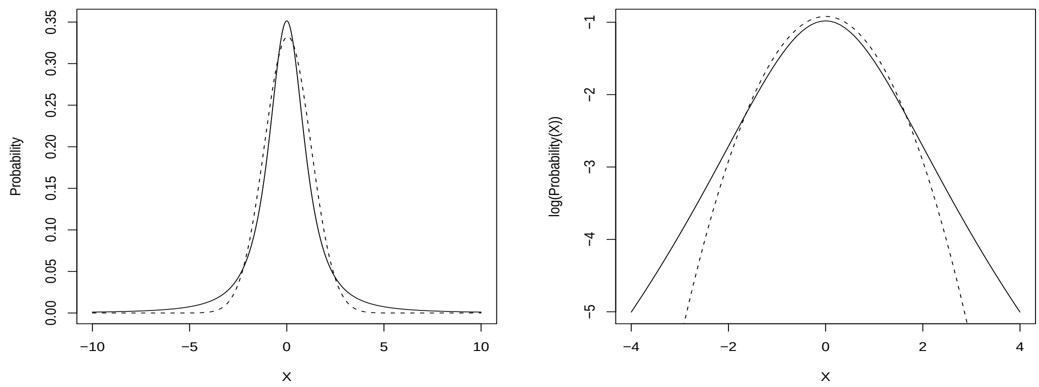


Figure 20: The charts show a Student-t distribution with an excess kurtosis of 6.7 (solid) in comparison with a normal distribution (doted). The left chart uses a linear y-axis whereby the right chart uses a logarithmic y-axis to make the excess kurtosis more explicitly. In the log chart appear the the fat tails of the student distribution as a line above the tails of the normal distribution.

With these definitions, the normal distribution has a skewness and a kurtosis of 0. The Jarque-Bera test calculates the skewness and kurtosis of a given distribution to find out, whether it is a normal distribution (with a value of 0 for both) or not. The test statistics is as follows:

If we assume normality for the underlying distribution, the standard error for the estimated skewness $\hat{\varsigma}$ and kurtosis $\hat{\kappa}$ are approximately $\sqrt{\frac{6}{N}}$ and $\sqrt{\frac{24}{N}}$ with N as the sample size. The Jarque-Bera test is defined as

$$JB = N[(\frac{\hat{\varsigma}^2}{6}) + (\frac{\hat{\kappa}^2}{24})] \tag{20}$$

and is asymptotically chi-squared with 2 degrees of freedom.

Quantile-Quantile plot

In this section we would like to present a graphical method to assign some sample data to a possible distribution. An α quantiles is defined as x such that

$$P[X < x] = \alpha$$

The quantile-quantile plot (QQ plot) is a scatter plot with the quantiles of the given empirical distribution on the vertical axis and the quantiles of the theoretical distribution on the horizontal axis. In order to calculate the quantiles of the empirical distribution, one first has to transform the empirical distribution according to the standard normal transformation (19). Now one can draw the QQ plot as scatter plot of the transformed empirical and the standard normal quantiles.

In [19] the main merits of a QQ plot are described as:

- If a random sample set is compared to its own distribution, the plot should look roughly linear.
- If there are a few outliers contained in the data, it is possible to identify them by looking at the scatter plot.
- If one distribution is transformed by a linear function, this transforms the QQ plot by the same linear transformation. The transformation can be estimated from the plot (slope and intercept)
- It is possible to deduce small differences in the participating distributions from the plot (e.g. fat tails imply curves at the left and right end)

Figure 21 shows a QQ plot for a sample from a student-t distribution with excess kurtosis. A distribution with excess kurtosis has a larger probability for events with very large or very small values in comparison to the normal distribution. From this we can conclude that fat tails will appear in a QQ plot as deviation from the diagonal at the extreme values. The deviation will be upwards for the high values and downwards for the low values.

3.2 Dependencies Tests

Runs test for randomness

The runs test can be used to decide if a data set is from a random process. It uses the concept of a run which is defined as a sequence of increasing values or a sequence of decreasing values. The length of a run is defined as the number of values belonging to this run. The runs test is based on the binomial distribution, which defines the probability that the i-th value is larger or smaller than the $(i + 1)$-th value.

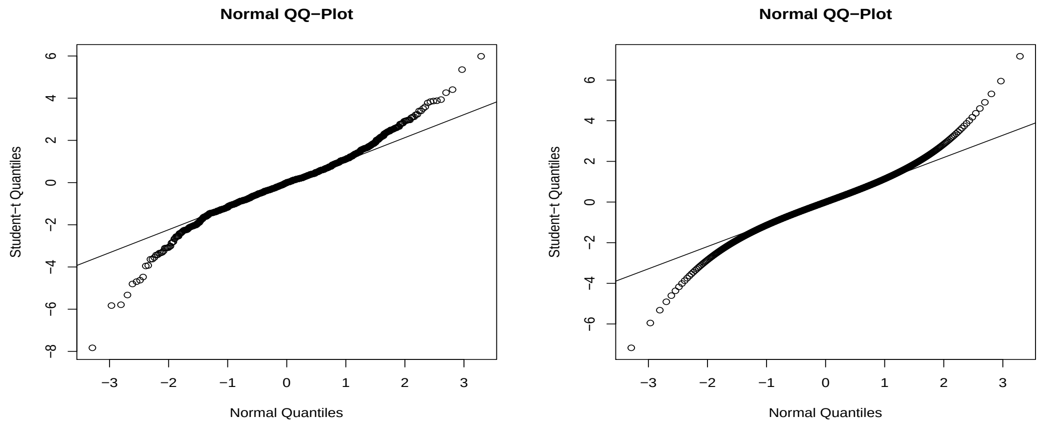


Figure 21: The charts show QQ plots for a student-t distribution with a degree of freedom of 4 in comparison with the normal distribution. The left chart was created from a sample set of 1000 elements from the student-t distribution and the right chart directly from the quantiles of the same normal and student-t distribution. The right chart is therefore smoother. The fat tails of the student distribution appear in both charts as deviation from the diagonal.

For the test we have to calculate the n_i's, the number of runs of length i for $1 \leq i \leq 10$. We can then normalize the n_i's with the expected number of runs of length i (μ_{n_i}) and the standard deviation of the number of runs of length i (σ_{n_i}). These values μ_{n_i} and σ_{n_i} can be received from the binomial distribution.
The final test value is the normalized n_i:

$$z_i = \frac{n_i - \mu_{n_i}}{\sigma_{n_i}}$$

which is compared to the two sided standard normal table. A z_i value greater than the table entry indicates non-randomness. Figure 22 and 23 show some outcome of AR(1) and GARCH(1,1) processes with the corresponding test results.

BDS test for dependencies

The BDS test is a non-parametric method of testing for nonlinear patterns in time series. It was first developed by Brock, Dechert and Scheinkman in 1987 (see [11]). The test has the null hypothesis that the data in the time series is independently and identically distributed (iid) and is in [8] defined as

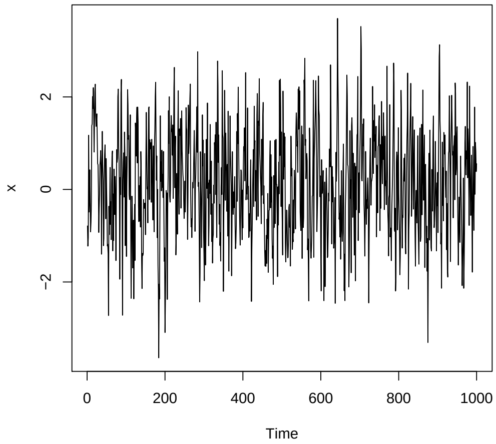
$$B_T = \frac{\sqrt{T-m+1}(C_T(m,\epsilon) - C_T(1,\epsilon)^m)}{\hat{\sigma}(m,\epsilon)}$$

where

- $C_T(m,\epsilon)$ is the correlation integral defined by

$$C_T(m\epsilon) = \binom{T-m}{2}^{-1} \sum_{\forall s<t} I_\epsilon(Y_t^m, Y_s^m)$$

- $Y_t^m = (y_t, y_{t+1}, \cdots, y_{t+m-1})$ is the m-history of y_t



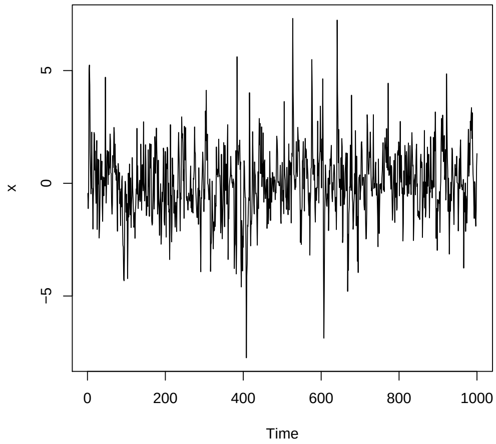


Figure 22: The charts shows a sample set derived from an AR(1) process with a coefficient $\phi = 0.5$. The process corresponding to the left picture had a standard normal distribution as innovation function and the process corresponding to the right picture had a student-t distribution with a degree of freedom of 4 for the innovation function. The Runs test calculates a value $n_1 = -0.70$ for the left chart and a value $n_1 = -0.55$ for the right chart. The standard normal table shows at the 5% significance level a value of 1.96. Since -0.70 and -0.55 is contained in ± 1.96, we can conclude that both underlying processes that have created the sample sets were random.

- $I_\epsilon(Y_t^m, Y_s^m)$ is the indicator function with $I_\epsilon(Y_t^m, Y_s^m) = 1$, if $\|Y_t^m, Y_s^m\| < \epsilon$, and $I_\epsilon(Y_t^m, Y_s^m) = 0$ otherwise. ϵ is a positive constant
- $\|Y_t^m, Y_s^m\|$ is the max-norm of Y_t^m, Y_s^m:

$$\|Y_t^m, Y_s^m\| := max(|y_t - y_s|, |y_{t+1} - y_{s+1}|, \ldots, |y_{t+m-1} - y_{s+m-1}|).$$

- $\hat{\sigma}^2(m, \epsilon)$ is a consistent estimator of the asymptotic variance of $\sqrt{T - m + 1} C_T(m, \epsilon)$

The underlying idea of the BDS test can be seen in the following:
The random event $\{I_\epsilon(Y_t^m, Y_s^m) = 1\}$ is the same as

$$\{\|Y_t^m, Y_s^m\| < \epsilon\} = \{|y_t, y_s| < \epsilon\} \cap \ldots \cap \{|y_{t+m-1}, y_{s+m-1}| < \epsilon\}$$

Let $A_{t,s}(m, \epsilon) = \{|y_t, y_s| < \epsilon\}$. The above relationship can be expressed as

$$A_{t,s}(m, \epsilon) = A_{t,s}(1, \epsilon) \cap \ldots \cap A_{t+m-1,s+m-1}(1, \epsilon)$$

If $\{y_t\}$ is an i.i.d. sequence, then the events $A_{t,s}(1, \epsilon), \ldots, A_{t+m-1,s+m-1}(1, \epsilon)$ will be independent, so

$$P[A_{t,s}(m, \epsilon)] = P[A_{t,s}(1, \epsilon)]^m$$

Since the correlation integral $C_T(m, \epsilon)$ converges in distribution to $P[A_{t,s}(1, \epsilon)]^m$, the BDS test detects the null hypothesis of serial independence by comparing if $C_T(m, \epsilon)$ is sufficiently close to $C_T(1, \epsilon)^m$.
The BDS statistic is easy to compute, however it has a disadvantage: The user has to define the two free parameters maximum embedding dimension m and relative radius ϵ *ex ante*.

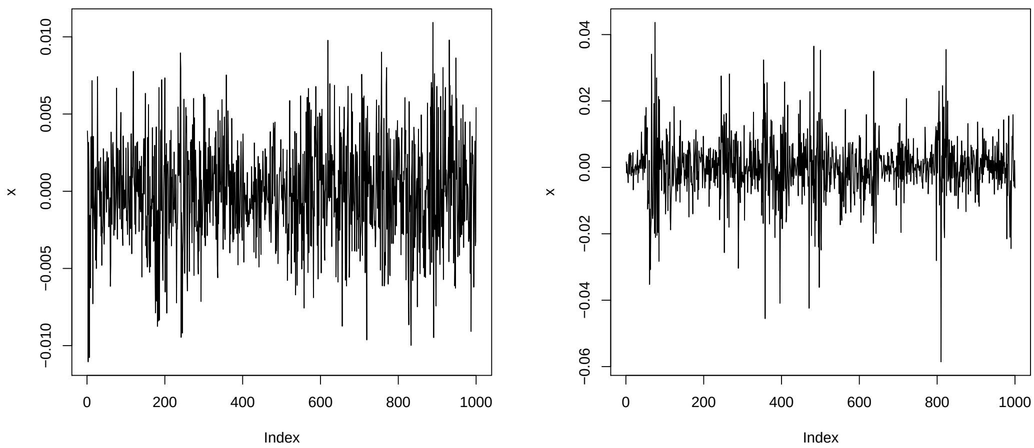


Figure 23: The charts show the same calculations as in figure 22 with an GARCH process (as described in [10]) as underlying function. Again, the process corresponding to the left picture had a standard normal distribution as innovation function and the process corresponding to the right picture had a student-t distribution with a degree of freedom of 4 for the innovation function. The Runs test calculates a value $n_1 = -0.65$ for the left chart and a value $n_1 = -0.62$ for the right chart. So we can again conclude that both underlying processes that have created the sample sets were random.

We will use the same AR(1) and GARCH(1,1) processes as described in figure 22 and 23 for the Runs test and apply the BDS test to them. The following part shows the detailed BDS analysis for the AR(1) process with normal innovation:

Embedding dimension = 2, 3
Epsilon for close points = 0.5836, 1.1672, 1.7508, 2.3344
Standard Normal =

	[0.5836]	[1.1672]	[1.7508]	[2.3344]
2	15.7714	16.6025	17.0971	18.2606
3	14.0135	14.8726	15.2300	16.4728

p-value =

	[0.5836]	[1.1672]	[1.7508]	[2.3344]
2	0	0	0	0
3	0	0	0	0

The test program has decided to use 0.58, 1.2, 1.8 and 2.3 as ϵ and calculate the statistics for embedding dimension 2 and 3. The first table shows the test results for each combination if embedding dimension and ϵ. Since all values lie above the threshold given by the standard normal distribution, we can (correctly) conclude that the series is not independent. The second table shows the p-values for the statistics. We can have great confidence in the results because of the very low p-values.
The following table summarizes the results of the BDS test applied to the four processes:

Process	Innovation Function	used ϵ	range of test results
AR(1)	Standard Normal	0.58 1.2 1.8 2.3	14 - 18
	Student-t	0.81 1.6 2.4 3.2	14 - 19
GARCH	Standard Normal	0.0016 0.0032 0.0049 0.0065	2.8 - 4.9
	Student-t	0.0038 0.0076 0.011 0.015	9.2 - 14

The results for the AR(1) process with student-t distribution as innovation function lie between 14 and 19 for ϵ as 0.81 1.6 2.4 3.2 and is therefore also not produced by and independent process. In the case of the GARCH process with normal innovation function as underlying function the results are not so ambiguous. We get values between 2.8 and 4.9 for the test statistics which is still larger than the corresponding value for the standard normal distribution and therefore we can also this time series declare as not independent. The reason for these small values might lie in the fact that the test programm has chosen the relative radius ϵ very small: 0.0016, 0.0032, 0.0049 and 0.0065.
A BDS test for GARCH with student innovation function produces values between 9.2 and 14 as test statistics. The ϵ is chosen as 0.0038, 0.0076, 0.011 and 0.015. Therefore we can conclude that this time series is also not independent.

3.3 Results Of Statistical Tests Applied To Market Data

Kolmogorov-Smirnov test

First we apply the market data to the Kolmogorov-Smirnov test to get an impression about whether they are normally distributed. We have calculated the test results for all of the listed market time series. The Smirnov-Kolmogorov test value is determined for different data intervals. This means that the given daily data (D) was aggregated to bi-daily data (BD), weekly data (W), bi-weekly data (BW), monthly data (M) and quarterly data (Q). For each of this data set and each mentioned index the test result is calculated. The values are listed in the following table. Each column represents an index, whereby E stands for 'Equity' and B for 'Bond'. Each row contains a time interval, abbreviated as explained above.

Interval	E World	E EU	E US	E FE	E CH	B World	B EU	B US	B FE	B CH
D	4.6	4.3	3.8	4.6	4.4	4.4	4.6	3.7	3.9	4.5
BD	4.3	4.3	3.6	3.5	3.9	4.0	3.5	3.4	3.5	3.7
W	3.8	3.2	3.2	3.8	3.8	3.2	3.8	3.2	3.6	3.7
BW	4.6	3.5	3.3	3.7	3.7	2.7	3.5	3.4	4.1	3.0
M	3.2	2.5	3.1	2.4	3.8	2.8	2.6	3.5	3.0	2.7
Q	3.3	2.0	2.4	2.7	2.2	3.1	2.4	3.1	4.0	2.4

According to the results, no time series is assumed to be normally distributed. However we can see, that the lower the data frequency, the closer we get to the confidence value and therefore to normally distributed returns.

Skewness, Kurtosis and Jarque-Bera test

We have calculated the skewness and kurtosis for all of the listed market time series except MSCI Europe and Lehman Aggregated Euro Bond Index since there is too few data available for these two indices. For all of the others we have taken the last 1953 samples points of the available data, i.e. all available data from the SBI Foreigner index and the last 1953 samples from some of the other used indices. Again we have aggregated the daily data to get also lower frequency data. The values for the skewness are listed in the following table.

Interval	E World	E US	E FE	E CH	B World	B US	B FE	B CH
D	-0.11	-0.089	0.32	-0.094	0.17	-0.47	-0.46	0.32
BD	-0.23	-0.046	0.30	-0.22	0.063	-0.47	-0.39	0.30
W	-0.40	-0.35	0.59	-0.57	-0.073	-0.50	-0.061	0.20
BW	-0.38	-0.50	0.94	-0.61	0.25	-0.48	-0.44	0.43
M	-0.54	-0.27	0.60	-0.66	0.72	-0.34	-1.0	0.60
Q	0.29	0.050	0.030	0.10	0.50	-0.090	0.26	0.65

The next table shows the respective values for the kurtosis.

Interval	E World	E US	E FE	E CH	B World	B US	B FE	B CH
D	5.0	5.5	6.6	5.6	4.8	5.5	9.2	5.2
BD	4.3	3.9	5.4	6.6	3.6	4.7	7.5	3.9
W	3.6	4.5	5.0	4.7	3.4	4.1	4.9	3.4
BW	5.1	5.2	5.5	7.4	3.4	3.7	4.0	3.0
M	3.6	2.6	3.6	4.2	4.1	3.1	6.5	3.5
Q	2.6	2.4	2.3	3.1	2.5	2.4	3.7	3.2

The results for the kurtosis are also summarized in figure 24. It is visible that the value for the kurtosis tends, for longer data periods, towards the value of the kurtosis of the normal distribution, which is 3. From this we can conclude that time series with a longer time interval like monthly or quarterly data can be better fitted to a normal distribution than data with higher frequency like intra-day or daily data which exhibits excess kurtosis. In [5], page 287, we can also find the conclusion that in most liquid financial markets is highly significant excess kurtosis in intra-day returns, which decreases with sampling frequency.

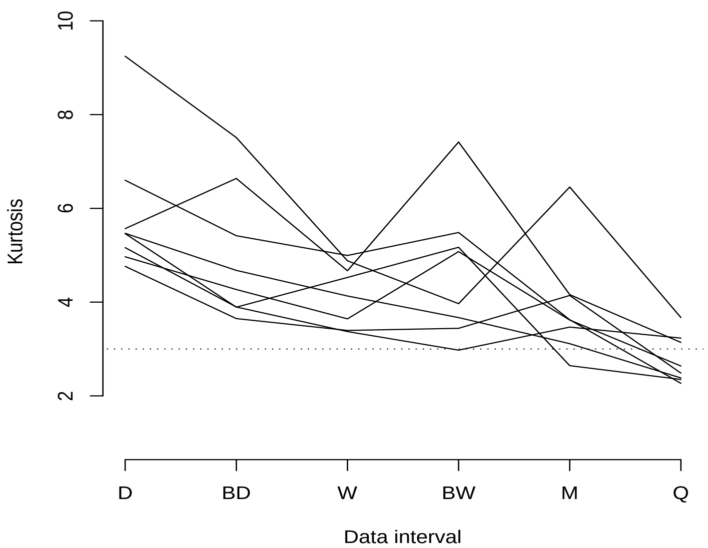


Figure 24: The chart shows the evolution of the kurtosis for several market time series and data intervals. Each line depicts a certain market time series for increasing interval lengths. The length of an interval is encoded according to: D: daily, BD: bi-daily, W: weekly, BW: bi-weekly, M: monthly, Q: quarterly. We can see that, for longer the data intervals, the values for the kurtosis approach the kurtosis of a normal distribution (doted line).

According to (20) we can calculate the test statistics for the Jarque-Bera test out of the skewness and kurtosis. The resulting table looks like

Interval	E World	E US	E FE	E CH	B World	B US	B FE	B CH
D	320	500	1100	540	262	570	3200	410
BD	74	33	250	550	18	150	850	48
W	17	46	87	66	2.9	37	58	4.7
BW	40	46	79	170	3.6	11	14	6.0
M	6.0	1.6	7.0	12	13	1.8	63	6.4
Q	0.57	0.54	0.67	0.075	1.6	0.52	0.90	2.2

These results get compared with a χ^2 distribution with two degrees of freedom. This distribution has the threshold for a 5% confidence level at 5.99. From this we can conclude that the data are normal on a quarterly basis and for Equity World, Equity US and Bond Far East also on a monthly basis. For any shorter time interval the normality assumption does not hold.

QQ Plot

On the following page we have depicted some QQ plots for the time series of Equities World, Equities US, Equities Switzerland, Bonds World and Bonds US. The QQ plots of the same time series are on the same horizontal line, ordered from daily data, bi-daily data, weekly data to bi-weekly data. It is visible that the fat tails disappear with lower data frequency and the empirical line approaches the linear line. Further one can see that the bond returns have are less fat tailed that the equity returns. Another interesting phenomenon is that, specially on a weekly basis, the lower fat tails are stronger evolved than the upper tails. The reason might be that a crash occurs in a shorter time interval than an euphoria.

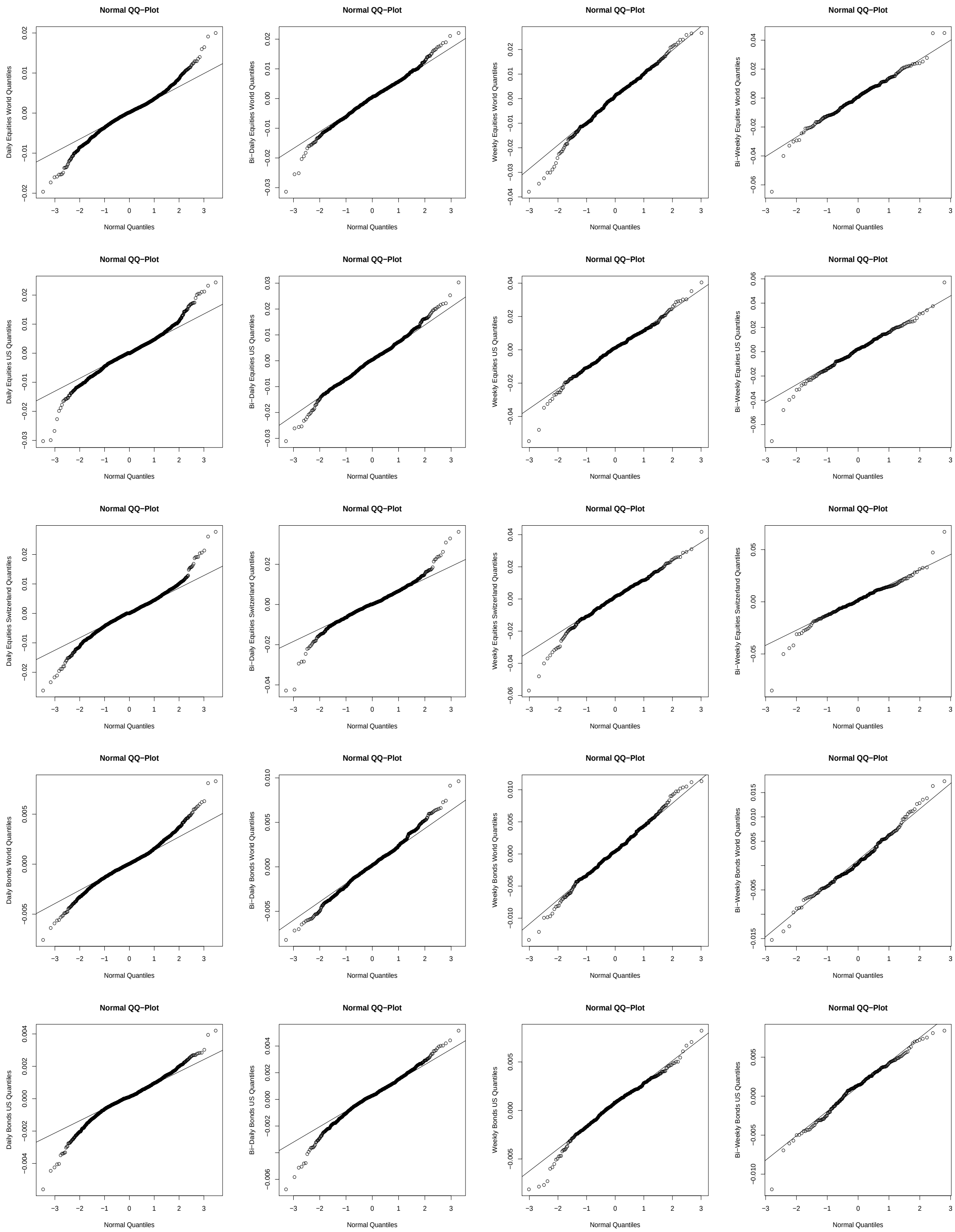


Figure 25: QQ Plots for Equities World, Equities US, Equities Switzerland, Bonds World and Bonds US time series and data intervals of daily data, bi-daily data, weekly data and bi-weekly data.

Runs test

In the following we show the results of the Runs test applied to the market data. The first table contains the result of the equity indices:

Interval	Equities World	Equities EU	Equities US	Equities FE	Equities CH
D	-0.92	-0.68	-0.60	-0.65	-0.67
BD	-0.78	-0.64	-0.64	-0.64	-0.70
W	-0.58	-0.49	-0.52	-0.60	-0.56
BW	-0.75	-0.73	-0.75	-0.58	-0.66
M	-0.76	-0.54	-0.61	-0.72	-0.68
Q	-0.36	-0.79	-0.65	-1.6	-0.36

A two sided standard normal distribution table gives us a value of 1.96 for the 5% significance level. Since all results are smaller than this threshold, we have to conclude that they are all generated by a random process.

Interval	Bonds World	Bonds EU	Bonds US	Bonds FE	Bonds CH
D	-0.68	-0.74	-0.79	-0.64	-0.62
BD	-0.67	-0.80	-0.70	-0.63	-0.55
W	-0.56	-0.86	-0.78	-0.83	-0.53
BW	-0.61	-1.07	-1.07	-0.85	-0.55
M	-1.1	-1.2	-1.0	-1.4	-0.72
Q	-1.1	-1.5	-6.6	-0.78	-0.78

The same hold true for the bonds indices because they also lie all in between the boundaries. We can find that the bond indices have smaller values and are therefore more likely to be randomly distributed.

BDS test

Finally, let's have a look at the results of the BDS test. We list the range of the test values for different values for the ϵ and embedding dimension.

Interval	Equities World	Equities EU	Equities US	Equities FE	Equities CH
D	6.9 - 13	8.9 - 44	2.8 - 8.9	2.1 - 6.2	7.6 - 11
BD	5.9 - 10	12 - 38	4.1 - 6.7	3.3 - 6.8	6.0 - 8.4
W	2.3 - 5.2	5.3 - 19	0.52 - 4.7	0.80 - 2.9	4.8 - 8.6
BW	-0.34 - 2.6	1.1 - 14	-0.62 - 2.0	-0.64 - 0.95	2.8 - 6.8
M	0.61 - 3.0	1.5 - 8.9	2.1 - 3.9	-2.2 - -0.45	-0.70 - 0.52
Q	-3.2 - 0.22	-5.3 - 0.33	-5.5 - 1.9	0.56 - 5.4	-1.8 - 1.7

The statement of the test (threshold 1.96) is that the market series are uncorrelated for monthly and quarterly data (except for the case of monthly data for Equities EU and US) and correlated for higher frequency data. Please remember that the results for Equities EU and Bonds EU are gained from a shorter time series than the others and are therefore not significant. The next table lists the range of the results for the bond indices:

Interval	Bonds World	Bonds EU	Bonds US	Bonds FE	Bonds CH
D	1.8 - 4.0	2.7 - 29	3.6 - 5.5	8.9 - 14	-0.27 - 1.44
BD	3.3 - 5.5	1.1 - 17	1.4 - 3.1	5.8 - 11	-1.1 - -0.51
W	2.1 - 3.8	0.021 - 5.7	2.8 - 3.7	3.7 - 10	-1.2 - 0.12
BW	0.26 - 2.6	-1.6 - 3.1	0.056 - 2.2	2.2 - 9.7	-1 - 1.0
M	-0.89 - 0.76	0.028 - 6.4	-2.6 - 0.23	0.61 - 6.1	-1.4 - 0.88
Q	-4.2 - 2.0	-1.2 - 8.3	-1.6 - 4.9	0.78 - 3.3	-3.7 - -1.3

We have more or less an acceptance of the hypothesis of uncorrelated returns for bi-weekly, monthly and quarterly data (except for bi-weekly Bonds Far East).

Summary of test results

Let's summarize the results of the applied tests:

- The Kolmogorov-Smirnov test has shown that the market time series are not normally distributed, neither on short time frequency (daily data), nor on long time frequency (quarterly data).
- The Jarque-Bera test confirms these statement by refusing normality except for quarterly data.
- The QQ plots for different sampling frequencies of the data show significant fat tails for daily up to bi-weekly data
- From the Runs test we were able to conclude that the market series were produced by a random process
- Finally, the BDS test showed us that the series are uncorrelated for monthly and quarterly data and correlated for higher frequency data

These results should be evidence enough that the normality assumption of Markowitz does not hold and it is justified to look out for other approaches that take non-normality into consideration. It was even proposed by Markowitz himself in his Nobel price winning work, also to investigate alternatives to variance as risk measures. There are some arguments for the standard Markowitz method which we don't want to hide:
The Central Limit Theorem says: Let $X_1, X_2, \ldots, X_n$ be mutually independent random variables with a common distribution function F. Assume E[X]= 0 and Var(X)= 1. As $n \to \infty$ the distribution of the normalized sum

$$S_n = \frac{(X_1 + X_2 + \ldots + X_n)}{\sqrt{n}}$$

tends to the Gaussian distribution function. When we look at the tick-by-tick logarithmic return data of a stock exchange for a certain financial instrument, we can interpret each data point as the value of a random variable and the daily, weekly or monthly data of this instrument as the sum of the tick-by-tick returns or the respective random variables. According to the Central Limit Theorem, the low frequency data will distribute like a Gaussian distribution function, if the frequency is low enough and we have enough data points in a period. In the context of an index or fund, the Central Limit Theorem can be applied once more by arguing that an index or fund is the weighted sum of several random variable (the constituents of the index or fund) and therefore the returns will behave according to a normal distribution if the index or fund has enough constituents.

4 Portfolio Construction With Non Normal Asset Returns

The concept of a mean risk framework was explained an earlier chapter. Markowitz uses this framework and has the variance chosen as risk measure. We will explore what general properties such a risk measure should have in order to be an substitute for the variance. In the second part of this chapter the suitability of variance as risk measure gets analyzed.

4.1 Introduction To Risk In General

In this section we will concentrate on the properties of financial risk measures. Part of the basic theory for this area was developed for the insurance sector and then adapted for the financial context.

We will use a variable X as a random variable representing the relative or absolute return of an asset (or the insured losses in the insurance context). Assume we have two alternatives A and B and their financial consequences X_A and X_B. Let the function R denote a risk measure which assigns a value to each alternative and the notation $A \succ_R B \Leftrightarrow R(X_A) > R(X_B)$ indicates that the alternative A is riskier then alternative B. Note that this is different from the utility function U presented in chapter 1 where $A \succ B \Leftrightarrow U(X_A) > U(X_B)$ means that A is preferred to B.
The difference of the concepts of the utility function and risk might become more apparent if one becomes aware that a utility function can be defined without a risk term (e.g. 'prefer more to less') or can include a risk term (e.g. Markowitz approach where we can find a trade-off function between expected return and risk).

In Albrecht [4] risk measures are categorized into two kinds. The two categories are:
1.) Risk as magnitude of deviation from target (risk of the first kind)
2.) Risk as necessary capital respectively necessary premium (risk of the second kind)
For many common risk measures one kind can get transformed into the other: The addition of E[X] to a risk measure of the second kind will guide us to a risk measure of the first kind and the subtraction of E[X] from a risk measure of the first kind will lead to a risk measure of the second kind.

We can find a general approach to derive a risk measure for a given utility function. This standard measure of risk is given in Huerlimann [26] by:

$$R(X) = -E[U(X - E[X])] \tag{21}$$

Since the risk measure corresponds to the negative expected utility function of X - E[X], the risk measure is location free. From (21) we can derive specific risk measures by using a specific utility function. Using for instance the quadratic utility function (6), we obtain the variance

$$Var(X) = E[(X - E[X])^2]$$

as the corresponding risk measure.

We will now introduce the definitions for stochastic and monotonic dominance because they are useful in the context of risk measures. Assume we are given two random variables X, Y.

Stochastic dominance of order 1 for a monotonic function R:

$$X \prec_{SD(1)} Y \leftrightarrow E[R(X)] \leq E[R(Y)]$$

Stochastic dominance of order 2 for a concave, monotonic function R:

$$X \prec_{SD(2)} Y \leftrightarrow E[R(X)] \leq E[R(Y)]$$

Monotonic dominance of order 2 for a concave function R:

$$X \prec_{MD(1)} Y \leftrightarrow E[R(X)] \leq E[R(Y)]$$

Next we will now check some axiomatic systems for risk measures that were proposed in the last years.

Axiomatic system of Pedersen and Satchell

Pedersen and Satchell give in [32] the following set of axioms for a risk measure:
1.) Nonnegativity: $R(X) \geq 0$
This requirement follows from the assumption of a risk measure of the first kind (deviation from a location measure)

2.) Positive homogeneity: $R(c * X) = c * R(X), \forall$ constants c
If an investment gets multiplied, then also the risk gets multiplied.

3.) Subadditivity: $R(X + Y) \leq R(X) + R(Y)$
The risk or two combined investments will not be larger than the risk of the individual investments (effect of diversification).

4.) Shift invariance: $R(X + c) \leq R(X), \forall$ constants c
The measure is invariant to an addition of a constant (location free)

Axioms number 2 and 3 combined lead to the statement that the risk of a constant random variable must be zero. Axioms 2 and 4 imply that a risk measure according to these criteria is convex. Since the risk measure is assumed to be location free, this system of axioms will describe especially risk measures of the first kind.

Axiomatic system of Artzner, Delbaen, Eber and Heath

Artzner, Delbaen, Eber and Heath [7] have developed another set of axioms. Risk measures that fulfill their properties are called coherent. The classification was refined in [13] to introduce the terms convex risk measure. Axioms 1 and 4 are also contained in the set of Pedersen and Satchell [32] in a similar way.
They call a mapping a convex risk measure if $\forall X, Y \in \Re^{\infty}$,

1.) Subadditivity: $R(X + Y) \leq R(X) + R(Y)$

2.) Monotonicity: $X \leq Y \Rightarrow R(X) \geq R(Y)$
A higher loss potential (statistical dominance) implies a higher risk.

3.) Translation Invariance: $R(X+a) = R(X) - a, \forall$ constant returns a
There is no additional risk for an investment without uncertainty.

A convex risk measure R is called a coherent risk measure if it satisfy the additional property:

4.) Positive homogeneity: $R(c * X) = c * R(X), \forall$ constant c

This set of risk axioms is well suited for risk measures of the second kind. In fact, every reasonable risk measure must be convex because a risk measure that does not satisfy subadditivity penalizes diversification and would not assign risk in an intuitive way.

Axiomatic system of Wang, Young and Panjier

Another important set of risk axioms was introduced by Wang, Young and Panjier [45]. They are dealing with premia in an insurance context, which can however easily be transferred to the financial context. The two main tasks in insurance markets are the calculation of the risk premia and the risk capital. The closed system of axioms for premia by Wang, Young and Panjier asks for some continuity properties and

1.) Monotonicity: $X \leq Y \Rightarrow R(Y) \leq R(X)$

2.) Comonotone additivity: X_1, X_2 comonotone $\Rightarrow R(X+Y) = R(X) + R(Y)$

Comonotone: $\exists$ random variable Z and monotone functions f and g with $X = f(Z)$ and $Y = g(Z)$

A general risk measure

Stone [42] reports a general risk measure containing the three parameters c, k and z as:

$$R(X) = [\int_{\infty}^{z} (|x-c|)^k f(x)dx]^{\frac{1}{k}}$$

The standard deviation and semi-standard deviation are part of this general risk measure class. This class was extended in [32] to a five parameter model:

$$R(X) = [\int_{\infty}^{z} (|x-c|)^a w[F(x)] f(x)dx]^b$$

which contains also the variance, the semi-variance and some other risk measures.

4.2 Variance As Risk Measure

Variance was proposed as appropriate measures for risk by Markowitz in his approach (7). The advantage of variance as risk dimension is that it is a very convenient and intuitive measure. It is very common in statistics and has for this reason well known properties. However it has also some properties that makes it not optimal as risk measure for financial applications.
The risk of very rare events are not taken into account very well by variance. We will show with the tests presented before that the returns of financial assets often have fat tails. This means that extreme events (very high returns or very high losses) are more likely than compared to a

normal distribution. In practice of portfolio optimization it is crucial to avoid very high losses because a lot of clients just ask for a preservation of their wealth. It is true that the variance penalizes extreme events by calculating the squared distance to the mean, however we should ask for something more specific. For this reason a risk measure that does not pay special attention for these kind of events is not very qualified.
Another unpleasant property of variance is its symmetry. When we talk about risk, we think of the risk for a loss. However variance measures also the "risk" of a gain, which is in fact something desired for an investor. This gives rise to asymmetrical risk measures which take only care for losses.
We have already mentioned that it is shown in [39] that variance is only compatible to the the concept of a utility function under the assumption of normally distributed returns or a quadratic utility function which is a very strong restriction.

5 Value At Risk Measures

In this chapter we will present a first alternative risk measure to the standard deviation. It is called Value at Risk and belongs to the quantile based risk measures.
There are efforts undertaken to introduce regulations to the financial industry to get a better control for the risk that is taken by its participants and also to help the companies to get a better overview for the risk they hold. This was also the topic of the Basel Committee on Banking Supervision where, as a conclusion, they recommend Value at Risk as an appropriate risk measure.

5.1 Value At Risk

We define Value at Risk as:
Let $\alpha \in (0,1)$ be a given probability level and w the asset weights of a portfolio. The Value at Risk at level α for the return R is defined as

$$VaR_\alpha(R_P) = sup\{x|P[R_P < x] \leq \alpha\} = F_{R_P}^{-1}(\alpha) \tag{22}$$

The function $F_{R_P}^{-1}(\alpha)$ is called the generalized inverse of the cumulative distribution function $F_{R_P}(x) = P[R \leq x]$ of R_P and gives the α-quantile of R_P.
$VaR_\alpha(R_P)$ can be interpreted as the loss of a Portfolio that will be exceeded in only α*100 percent of all cases. Since α is usually chosen between 0.01 and 0.1, the Value at Risk is a lower boundary for a portfolio return and the return of the portfolio will with a very high probability (0.99 or 0.9 for the example α) not be smaller. It is the aim of portfolio construction to assemble a portfolio with a high Value at Risk in order to shift the return range for the $1-\alpha$ area as much to the positive side as possible. Sometimes α is chosen as 0.95 or 0.99 and $VaR_{1-\alpha}$ for a loss function is computed. The similarity of these two notations is shown in appendix B. Figure 26 shows two areas α and $1-\alpha$ for the normal distribution.

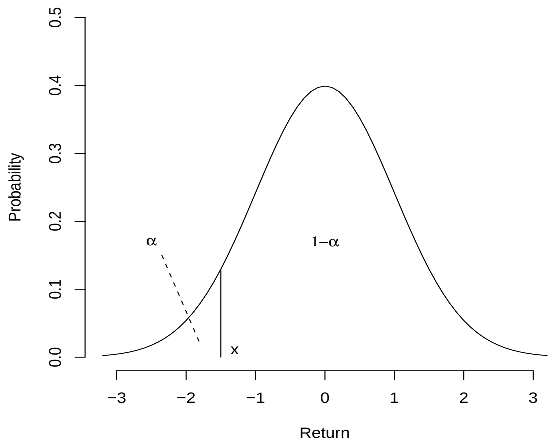


Figure 26: The Value at Risk at level α for the return R is defined as the return x where the probability of having a return smaller than x is α.

The analytical properties of the Value at Risk model are not very pleasant: It is in the general case not possibly to find an symbolic expression for the portfolio weight w optimized according to VaR and dependent on the multivariate returns function of its constituents. Even the numerical application is difficult: According to Gaivoronski and Pflug [23], VaR is not a convex risk measure. This means that the VaR function contains many local maxima. To deal with these maxima, they have developed a smoothing algorithm, which allows them to calculate the optimal portfolios in the VaR sense with high accuracy and in reasonable time. Another approach to deal with the VaR optimization function is proposed in Embrechts et al [20]. It treats each univariate distribution function for the assets individually and models the dependencies of the univariate distribution functions with a copula. The concept of the copula is a well known way of modelling dependence in risk management.

Another unpleasant property of Value at Risk is that it fails to be coherent as stated in [3]. In the general case Value at Risk does not fulfill the sub-additivity axiom. This is especially unpleasant because it implies that a portfolio made out of smaller portfolios (and therefore with a higher diversification as the individual small portfolio) can have a higher amount of risk than the sum of the risk of the smaller portfolios. This would offset the effect of diversification.

Since Value at Risk is only concerned about the threshold that will be crossed with the small probability α, it does not take into consideration the distribution of the returns above the threshold. Dembo and Fuma [17] published an example that shows this disadvantage. Assume two distributions are given as declared in this table and depicted in figure 27.

Return	Probability in Portfolio A	Probability in Portfolio B
-10	0.01	0.01
-7.5	0.04	0.04
-5	0.05	0.25
-2.5	0.1	0.25
0	0.5	0.3
2.5	0.15	0.1
5	0.15	0.05
μ	0.225	-1.775
σ	3.140	3.144
1% VaR	-10	-10
5% VaR	-7.5	-7.5

From the expected returns of the portfolio A and B we can see that Portfolio A has a higher expected return than portfolio B. This means that we have a clear preference for portfolio A. However both risk measures, standard deviation and VaR, fail to capture this preference because they both get the same values for portfolio A and B. The reason is that standard deviation, as mentioned, does not discriminate between the risk of a loss (which should be avoided) and the risk of a gain (which is favorable) and VaR does not take into consideration the distribution form above the threshold at all. This example is not as artificial as it might look like, since most distributions in finance differ especially around the median and are around the tails very similar. It is also clear that Value at Risk does not distinguish between very sever losses or just small losses, as long as they are below the threshold.

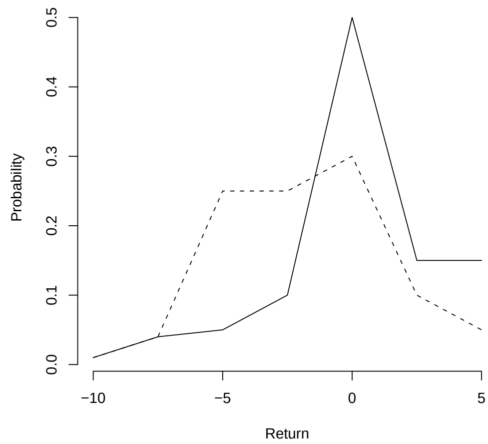


Figure 27: The graphic shows the two distribution functions defined in the table. The distribution function of the returns of portfolio A is depicted as solid line, whereas the distribution function of the returns of portfolio B is depicted as doted line. The functions are coincident for returns between -10 and -7.5

5.2 Conditional Value At Risk, Expected Shortfall And Tail Conditional Expectation

In this chapter we will discuss the concepts of Lower Partial Moments (LPM), Conditional Value at Risk (CVaR) and Expected Shortfall (ES). We cover them in the same chapter because they are very similar and these risk concepts have become a totum revolutum in the last few years. The following part tries to unveil the relation between the mentioned risk measures.

At the beginning there was a first concept called lower partial moment (LPM) as described in Fishburn [21]. The general lower partial moment risk measure for a random return variable R and its probability function $P(x)$ is given by

$$LPM_\beta(\tau; R) = E[(\tau - R)^\beta] = \int_{-\infty}^{\tau} P(x)(\tau - x)^\beta dx$$

An investor can determine a threshold τ under which he does not want the return R to fall. According to the choice of β, one gets a different lower partial moment:

$\beta = 0$: Shortfall probability $LPM_0 = \int_{-\infty}^{\tau} f(x)dx$

$\beta = 1$: Mean Shortfall $LPM_1 = \int_{-\infty}^{\tau} f(x)(\tau - x)dx$

$\beta = 2$: Shortfall variance/ Semi variance $LPM_2 = \int_{-\infty}^{\tau} f(x)(\tau - x)^2 dx$

LPM_0 portfolio selection corresponds to Roy's safety first rule presented in [38]. LPM_1, also called expected regret in [16], can be interpreted as the average portfolio underperformance compared to a fixed target or some benchmark τ.

The term conditional Value at Risk was first introduced in [35]. They use a slightly different definition and notation for VaR and CVaR as we will (refer to appendix B). For continuous distributions conditional Value at Risk is defined as conditional expected loss under the condition that it exceeds the Value at Risk. There are two variants of CVaR:

$$CVaR_{\alpha}^{+} = E[R_P | R_P < VaR] \tag{23}$$

$$CVaR_{\alpha}^{-} = E[R_P | R_P \leq VaR] \tag{24}$$

where $VaR = VaR_{\alpha}(R_P)$ as defined in formula (22) and $E[x]$ denotes the expected value of x.
As mentioned before, conditional Value at Risk can be considered as expected amount of loss below the VaR. From this it gets clear that $CVaR_{\alpha} \leq VaR_{\alpha}$

Conditional Value at Risk is also known as Mean Excess Loss (CVaR^{+}), Mean Shortfall (LPM$_1$ with $\tau = VaR$) (CVaR^{+}) or Tail Value at Risk (CVaR^{-}). Since the concept was developed for several application fields (e.g. actuarial science, finance, economics) and by different researchers, it has many names and definitions. In Huerlimann [25], ten equivalent definitions of CVaR are presented. A general definition for CVaR, also applicable for discrete distributions is written in Uryasev [44] as a weighted average of VaR and returns strictly below VaR. After the conversion to our environment the equation is

$$CVaR_{\alpha} = \lambda \ VaR_{\alpha} + (1 - \lambda) \ CVaR_{\alpha}^{+} \tag{25}$$

with

$$\lambda = \frac{\alpha - P[R_P \leq VaR]}{\alpha}$$

The equation can be used for continuous and discrete distributions: In the case of a continuous distribution $\lambda = 0$ and therefore $CVaR_{\alpha} = CVaR_{\alpha}^{+}$. If we have a discrete distribution the calculated VaR (VaR_{disc}) will not exactly be the α quantile as it would be for a continuous distribution (VaR_{cont}), but more on the negative side of the distribution. λ increases $CVaR_{\alpha}^{+}$ to the positive side of the distribution and extrapolates $CVaR_{\alpha}^{+}$ from VaR_{disc} to VaR_{cont}. In other words, $CVaR_{\alpha}^{+}$ and VaR get weighted proportionally to $\frac{VaR_{disc} - VaR_{cont}}{VaR_{cont}}$ and therefore $CVaR_{\alpha}^{+} \leq CVaR_{\alpha} \leq VaR_{\alpha}$.

A similar concept to CVaR is called expected shortfall. It was introduced in [1] and redefined later to be consistent with CVaR.

$$ES_{\alpha}(R_P) = -\frac{1}{\alpha}(E[R_P 1_{R_P \leq VaR}] - (P[R_P \leq VaR] - \alpha)VaR) \tag{26}$$

They show in Acerbi and Tasche [2] that it can also be expressed as

$$ES_{\alpha}(R_P) = -\frac{1}{\alpha}\int_0^{\alpha} inf[x | P[X \leq x] \geq a] da$$

In case that we have a non continuous distribution function, it might be that $P[R_P \leq x] > \alpha$. In contrast, for a continuous distribution function $P[R_P \leq x] = \alpha$ and then it can be seen that (23) is equivalent to (26).

To conclude we try to group the risk measures that have the same base concept. They all take the distribution function as input and process a number as representant for the risk the distribution function holds out of the distribution function.

Calculate a threshold the returns should not fall below	Calculate the expected return of the returns under a certain threshold
LPM_0 Value at Risk Shortfall risk	LPM_1 Conditional Value at Risk Expected shortfall Mean shortfall Expected regret Tail Value at Risk

It is shown in Testuri and Uryasev [43] that expected regret and CVaR is closely related. They also confirm the relation of CVaR with the other risk measures in the same row, at least for the case of continuous distribution functions.

The following table lists the properties of Value at Risk and conditional Value at Risk/ expected shortfall. The statements were taken from [34].

Property	VaR	CVaR
Translation equivariance	$\surd$	$\surd$
Positively homogeneous	$\surd$	$\surd$
Convexity	x	$\surd$
Stochastic dominance of order 1	$\surd$	$\surd$
Stochastic dominance of order 2	x	$\surd$
Monotonic dominance of order 2	x	$\surd$
Coherence	x	$\surd$

From this comparison it shows that conditional Value at Risk has much nicer properties than the standard Value at Risk. Since CVaR is convex with respect to portfolio positions, it is much easier to optimize than VaR which has a lot of local maxima. Coherence is a requirement for an intuitive risk measure (effect of diversification) and is also fulfilled only by CVaR.
Conditional Value at Risk gets presented an excellent tool for risk management and portfolio optimization because it can quantify risks beyond Value at Risk and is easier to optimize.
In [35] it is also stated that CVaR methodology is consistent with Mean-Variance methodology under normality assumption. This means that a CVaR maximal portfolio is also variance minimal for normal return distributions.

We will now focus on the optimization of the two risk measures following [34]. For the sake of consistency, we have again transformed the notations according to appendix B. $R = (R_1, \dots R_N)$ indicates a vector of random returns of asset classes $1 \dots N$. Let $w = (w_1, \dots w_N)$ be the weights of the investments in these asset classes. We try to maximize the risk measure under the constraint that the expected return $w^T R$ of the portfolio is equal to some predefined level μ. The VaR optimization problem can be stated as

$$\begin{aligned}
&\text{Maximize (in w) } VaR_\alpha(w^T R) \\
&\text{s.t.} \\
&w^T E[R] = \mu \\
&w^T 1 = 1 \\
&w \geq 0
\end{aligned}$$

and the CVaR respectively as

Maximize (in w) $CVaR_\alpha(w^T R)$
s.t.
$w^T E[R] = \mu$
$w^T 1 = 1$
$w \geq 0$

Note that, since our optimizer is only capable of minimizing a function but not of maximizing a function, we minimize in the implementation $-VaR_\alpha(w^T R)$ and $-CVaR_\alpha(w^T R)$.
The VaR optimization problem we will cover later. First, we transform the CVaR optimization problem in the following linear program with a dummy variable Z:

Maximize (in w and a) $a + \frac{1}{\alpha}E[Z]$
s.t.
$Z \geq w^T R - a$
$x^T E[R] = \mu$
$w^T 1 = 1$
$Z \geq 0$
$w \geq 0$

Since we have only linear constraints, we can be sure that the solution will be a singleton, a convex polyhedron or the solution does not exist.
In practice however we have mostly discrete variables (e.g. empirical data). For this reason we formulate the portfolio optimization problems in a discrete way.
A vector $R^i, i = 1, \dots, M$ indicates the returns of all asset classes for a certain time point 1,. . ., M. For the formulation we will use a notation $S_{[1:k]}(u^1, \dots, u^M)$ to denote the one element among $u^1, \dots, u^M$ which is the k-th smallest. The new definitions for VaR and CVaR are

$$VaR_\alpha(w^T R) = S_{[1:\lfloor \alpha M \rfloor]}(w^T R^1, \dots, w^T R^M)$$
$$CVaR_\alpha(w^T R) = \frac{1}{M} \sum_{w^T R^i \leq VaR_\alpha} w^T R^i$$

The discrete portfolio optimization problem for the VaR is a nonlinear, nonconvex program:

Maximize (in w) $S_{[1:\lfloor \alpha M \rfloor]}(w^T R^1, \dots, w^T R^M)$
s.t.
$w^T e = \mu$
$w^T 1 = 1$
$w \geq 0$

where $e = \frac{1}{M} \sum_{i=1}^{M} R^i$ denotes the expected return vector.
The discrete version of the CVaR is piecewise linear and may therefore be solved using an LP-solver. We formulate the problem like:

Maximize (in w, a, and z) $a + \frac{1}{\alpha M} \sum_{i=1}^{M} z^i$
s.t.
$z^i \geq -w^T R^i - a$
$w^T e = \mu$
$w^T 1 = 1$

$$z^i \geq 0$$
$$w_i \geq 0$$

For this setting, the optimal value for a is $VaR_\alpha(w^T R)$. We can see that the objective function and the first and third inequality constraint express the weighted average of the Value at Risk and the mean of all negative returns above the Value at Risk (which is the same as the mean of all returns below the negative Value at Risk).

5.3 Mean-Conditional Value At Risk Efficient Portfolios

In this section we want to analyze what it means to optimize a portfolio regarding Value At Risk/ Conditional Value At Risk.
We start with the case of normal distributed asset returns. Figure 28 shows two normal distribution with the same mean but different variances. The distribution with the larger variance (doted line) has smaller CVaR and vice-versa. It is intuitive to see that if we maximize the CVaR we also minimize the variance of the distribution. The only way to enlarge the CVaR (shifting the corresponding left tail of the distribution to the right) is to shorten the variance (make the peak larger). Of course this is only true if we a sufficient amount of data coming from a pure normal distribution function. Using small amounts of empirical data, there might be effects that prevent the equivalence of the two optimization techniques.

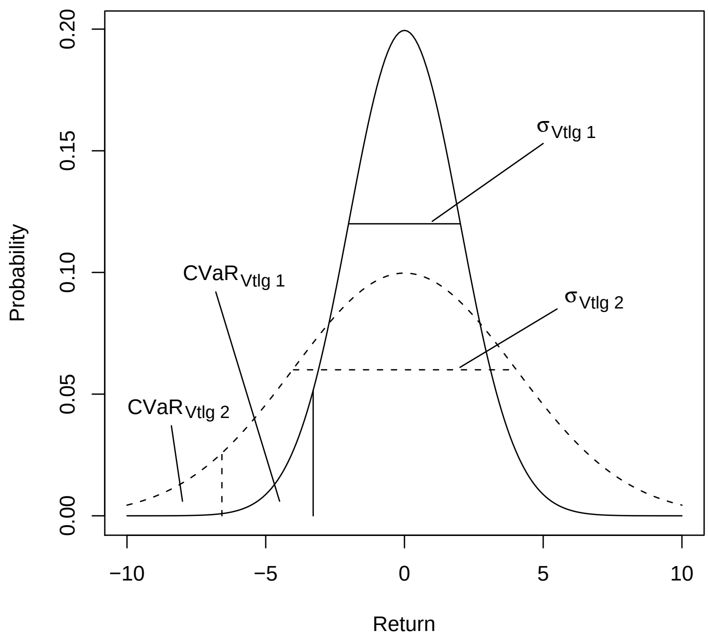


Figure 28: The graphic shows two normal distribution functions. For both distribution functions the Conditional Value at Risk and the variance is schematically depicted. It shows that minimizing the variance of a function is equivalent to maximizing its Conditional Value at Risk.

The case of distributions with skewness and excess kurtosis is more interesting. The occurrence of fat tails and asymmetry in the distribution function allows the mean-CVaR optimization to take the risk evolving out of these properties into account. As consequence, such an optimization will assign the portfolio weights differently the the Mean-Variance approach. The optimization, in general, will prefer assets with positive skewness, small kurtosis and low variance for a given return.

To conclude, we expect the results of a Mean-CVaR and the results of a Mean-Variance optimization to be the same for the case of similar distribution functions (e.g. normal distribution functions) for the asset returns and a sufficient amount of data. Au contraire, the results are assumed to be different for the two optimization techniques if the data is coming from varying distribution function with different higher moments or if the sample size is small.

6 Draw-Down Measures

In this section we will present two other approaches to measure the risk of a portfolio. They are called Draw-Down and Time Under-The-Water. Draw-Down was first presented in a portfolio context in [12]. In [33] Draw-Down is used together with Time Under-The-Water to measure the loss potential of hedge funds. We will describe Draw-Down as written in [12] and Time Under-The-Water according the idea in [33]. Afterwards we will enhance Draw-Down to Conditional Draw-Down at Risk (CDaR) and Time Under-The-Water to Conditional Time Under-The-Water at Risk (CTaR). Finally we apply CDaR in a portfolio context.

6.1 Draw-Down And Time Under-The-Water

An advantage of the two concepts is that they are much more intuitive than other risk measures. The concepts represent values every investor is interested in: Draw-Down measures the loss the investment might suffer (in absolute or relative terms) and Time Under-The-Water is the time period the investment might remain with a negative performance. Other possible applications for these measurements could be: A portfolio manager might loose a client if the clients portfolio does not provide a gain over a long time or a fund might not be allowed to loose more than a certain amount each month and has therefore to stop trading until the next month starts and therefore a new budget.

We will work on the logarithmic returns instead of geometric returns as stated in [12]. Assume we are given the (cumulated) return of the portfolio from time 0 until time t by a function

$$r_c(w,t)$$

with w as the vector of weights for the portfolio constituents. The Draw-Down function at time t is defined as the difference between the maximum of the function in the time period [0,t] (High-Water-Mark) and the value of the function at time t:

$$DD(w,t) = max_{0\leq\tau\leq t}[r_c(w,\tau)] - r_c(w,t) \tag{27}$$

Figure 29 shows a time series with the respective High-Water-Marks and Draw-Down.

Starting with the formula for Draw-Down, two risk functions are derived: Maximum Draw-Down is calculated as the maximum Draw-Down in the period

$$MD(w,t) = max_{0\leq\tau\leq t}[DD(w,t)] \tag{28}$$

and the average Draw-Down is defined as

$$AD(w) = \frac{1}{T}\int_0^T DD(w,t)dt \tag{29}$$

If, in a time-value framework, Draw-Down is measured on the y-axis, Time Under-The-Water is the corresponding period on the x-axis that represents the time the value of an investment may remain under its historic record mark. We define Time Under-The-Water as

$$TUW(w,t) = t - [maxT|r_c(w,maxT) = max_{0\leq\tau\leq t}r_c(w,\tau)]$$

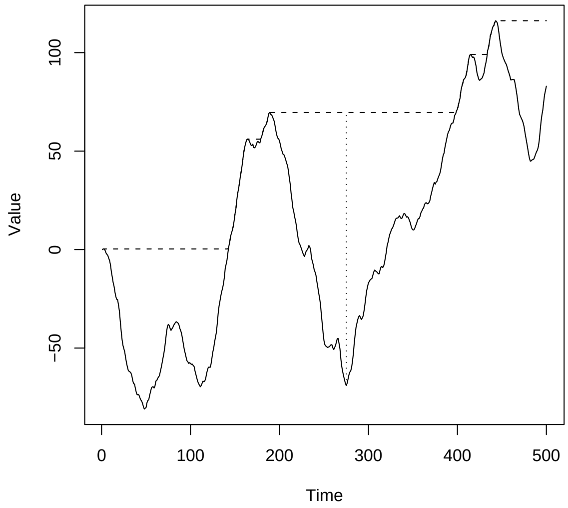


Figure 29: The figure shows a time series with the respective High-Water-Mark (dashed line) and Draw-Down (doted line) as defined. The Time Under-The-Water is just the part of the dashed line above the doted line.

Similar to the Draw-Down concept, we will now introduce Maximum Time Under-The-Water MT(w) and Average Time Under-The-Water AT(w) as

$$MT(w,t) = max_{0 \le \tau \le t}[TUW(w,t)] \tag{30}$$

$$AT(w) = \frac{1}{T}\int_0^T TUW(w,t)dt \tag{31}$$

6.2 Conditional Draw-Down At Risk And Conditional Time Under-The-Water At Risk

Alike the enhancement of Value at Risk to Conditional Value at Risk, we will proceed with Draw-Down and Time Under-The-Water. Draw-Down at Risk can be defined similar to (22) as

$$DaR_\alpha(MD) = inf\{x|P[MD > x] \le \alpha\} \tag{32}$$

with MD as Maximum Draw-Down and Conditional Draw-Down at Risk corresponding to Conditional Value at Risk (25) as:

$$CDaR_\alpha = \lambda \ DaR_\alpha + (1-\lambda) \ CDaR_\alpha^+ \tag{33}$$

with

$$\lambda = \frac{P[MD \ge DaR_{alpha}] - \alpha}{\alpha}$$

$$CDaR_\alpha^+ = E[MD|MD > DaR_\alpha] \tag{34}$$

We will now discuss the implementation of the concepts in detail. A first approach would be to calculate the Maximum Draw-Down for each new level of the High-Water-Mark. This means, we scan the time series from the past to the present and each time we find a new global maxima, we calculate the Maximum Draw-Down for the period between this global maxima and the point where the time series is higher than this global maxima for the first time. All of these Maximum Draw-Downs get stored to construct their distribution. The drawback of this method is that we will probably get very few Draw-Down values for the following reasons:

- Since the Draw-Down gets calculated as the difference to the highest historical value (record), the concept of the Draw-Down comprises the effect of increasing time periods for new records: The expected time period for a random variable to reach a new all-time-high is not uniformly distributed but increases much faster over time (see for example [19]).
- In times of a Baise, we won't get any Draw-Downs at all. Only in times of a Hausse there will be a new High-Water-Mark and therefore new Draw-Downs.

In [12] it was proposed to introduce M sub-periods in the time interval [0,T] and to calculate the Draw-Down for each sub-period. This way they get an empirical distribution for the Draw-Downs consisting out of maximum M sample points. Using this methodology, one should be aware of some points:

- The methodology adds a new variable M that does not improve the descriptive power of the concept. The reason for introducing this variable is just for numerical reasons and has no economical or practical meaning.
- If M is chosen too large, the number of resulting Draw-Downs is too small to get a good distribution approximation. If M is chosen too small, the Draw-Downs that extend over several sub-periods get cut into several smaller Draw-Downs because the maximum possible Draw-Down is restricted to the length of the sub-period. This is especially undesirable since we are particularly interested in the large Draw-Downs to calculate the α-quantile.
- The effect of increasing time periods for new records can not be avoided by resetting the all-time-high at the beginning of each sub-period - it is just transformed to a smaller time scale.

We would like to bring this method and a new method into the context of the information given by the client. The described method of the fixed periods for calculating the Draw-Down could be used if the investment horizon of the client is known:

If the investment horizon is known, a rolling window with the length of the investment horizon could be applied to the available historical data. For each time window the Maximum Draw-Down gets calculated and the window shifted for one period. If we have P historical data points, Q data points in the rolling window and shift the window R data points each time we advance, we get with this method $\frac{P-Q}{R}$ Maximum Draw-Down values. It would also be possible to use overlapping rolling windows. However this would decrease the variance of the in this way reused data.

If the investment horizon is not known, we propose as a second method to calculate the Maximum Draw-Down for each possible entry combined with each possible exit point. This gives us for P data points $\frac{(P-1)(P-2)}{2}$ Maximum Draw-Down values. The idea is to calculate the average Draw-Down an investor could face. The disadvantage of this method is that the Draw-Down values for a certain (unknown) investment horizon have very few influence to the final distribution. The reason for this is that the number of possible Draw-Downs grows quadratical, however the number of Draw-Downs for a certain investment period grows only linearly. It might therefore be questionable to compare Draw-Downs of different time periods.

We will assume that the investment horizon is known and therefore proceed with the first of the described methods to formulate the optimization problems.

It lies in the nature of the concept to change the structure of the optimization problem from "minimize the risk for a given expected return", as it was the case for Variance and CVaR optimization, to "maximize the expected return for a given Draw-Down/Time Under-The-Water threshold". For an investor it is convenient to define his/her personal amount of wealth he/she is willing to risk or the amount of time he/she gives to the portfolio manager to remain with a negative performance. However, to be better able to compare the results of the different optimizations, we will stick to our old schema of fixing an expected return and minimizing the respective risk measure.

To show the corresponding linear optimization problems, we introduce the following variables: The vector of logarithmic cumulative asset returns up to time moment k be y_k so we can calculate the cumulative portfolio return as $r_c(w, t = k) = y_k * w$. With the expected return given by the investor as μ, we get the following linear programming problem for the Maximum Draw-Dawn

$$
\begin{array}{ll}
\text{Minimize (in w and u) } z & \\
\text{s.t.} & \\
u_k - y_k * w \leq z & 1 \leq k \leq M \\
u_k \geq y_k * w, & 1 \leq k \leq M \\
u_k \geq u_{k-1}, & 1 \leq k \leq M \\
u_0 = 0 & \\
\frac{1}{d} y_M * w = \mu & \\
w^T 1 = 1 & \\
w_i \geq 0, & 1 \leq i \leq N
\end{array}
$$

where u_k , $1 \leq k \leq M$ and z are auxiliary variables and d is the investment period in years.

The optimization problem with a constraint on the average Draw-Down can be written as follows

$$
\begin{array}{ll}
\text{Minimize (in w and u) } z & \\
\text{s.t.} & \\
\frac{1}{M} \sum_{k=1}^{M} (u_k - y_k * x) \leq z & \\
u_k \geq y_k * x, & 1 \leq k \leq M \\
u_k \geq u_{k-1}, & 1 \leq k \leq M \\
u_0 = 0 & \\
\frac{1}{d} y_M * x & \\
w^T 1 = 1 & \\
w_i \geq 0, & 1 \leq i \leq N
\end{array}
$$

and the optimization problem with a constraint on CDaR may be formulated as

$$
\begin{array}{ll}
\text{Minimize (in w, u, z, } \zeta) \; z & \\
\text{s.t.} & \\
\zeta + \frac{1}{\alpha M} \sum_{k=1}^{M} z_k \leq z & \\
z_k \geq u_k - y_k * x - \zeta, & 1 \leq k \leq M \\
z_k \geq 0, & 1 \leq k \leq M \\
u_k \geq y_k * x, & 1 \leq k \leq M \\
u_k \geq u_{k-1}, & 1 \leq k \leq M \\
u_0 = 0 & \\
\frac{1}{d} y_T * x & \\
w^T 1 = 1 & \\
w_i \geq 0, & 1 \leq i \leq N
\end{array}
$$

The optimal solution of this problem gives the optimal threshold value in variable ζ.

The corresponding extension of the Time Under-The-Water to a risk measure Conditional Time Under-The-Water at Risk (CTaR) can be done similarly.

Since the optimization problems are analogous to the ones of Draw-Down, we will give only the linear programm for the Conditional Time Under-The-Water at Risk

$$
\begin{array}{ll}
\text{Minimize (in w, u, z, } \zeta) \; v & \\
\text{s.t.} & \\
\vartheta + \frac{1}{\alpha M} \sum_{k=1}^{M} z_k \leq v & \\
z_k \geq u_k - y_k * x - \vartheta, & 1 \leq k \leq M \\
z_k \geq 0, & 1 \leq k \leq M \\
u_k \geq y_k * x, & 1 \leq k \leq M \\
u_k \geq u_{k-1}, & 1 \leq k \leq M \\
\frac{1}{d} y_T * x & \\
u_0 = 0 & \\
w^M 1 = 1 & \\
w_i \geq 0, & 1 \leq i \leq N
\end{array}
$$

where u_k , z_k, $1 \leq k \leq M$ and v are auxiliary variables.

A well implementable setup for the portfolio optimization process (that we will not further follow) would be the following framework: Optimize the expected portfolio return subject to the clients CDaR restriction ζ_C and CTaR restriction η_C.

$$
\begin{array}{l}
\text{Maximize } \frac{1}{dC} y_M * x \\
\text{s.t.} \\
CDaR_\alpha(w) \leq \zeta_C \\
CTaR_\alpha(w) \leq \eta_C
\end{array}
$$

This setup has the following advantages:

- Since it is very intuitive, the portfolio manager can talk to the client in exactly the same terms and the client has a clear view about the risk he or she is taking.
- The framework is still a linear programming problem and can therefore be solved efficiently.

The two risk measures Draw-Down and Time Under-The-Water demand a lot of data to be meaningful. Both methods define at the most one risk value per data window and to get a good approximation for the distribution of the risk measures, lots of data windows are necessary. This is especially true when we are looking for small quantiles like $\alpha = 0.05$. In empirical test we have seen that Draw-Down and Time Under-The-Water as described so far are not very appropriate for hedge funds, where only monthly data for about the last 15 years is available and therefore only 180 data points in total. The risk measures are in this case to discrete and it is not possible to get a reasonable optimization. E.g. it is often not possible to get a good estimation for the derivatives of the risk measure which is needed in most optimization algorithms. We have come to this conclusion especially for the Time Under-The-Water measure where the objective function to minimize is far to discrete to get any meaningful results.

An important difference of the Draw-Down approach in comparison to the Value at Risk approach is the fact that the Draw-Down takes the correlations implied in the time series into consideration because it operates on the compounded historical returns and not on the return distribution function as Value at Risk does.

6.3 Mean-Conditional Draw-Down At Risk Efficient Portfolios

In this section we want again to analyze what it means to optimize a portfolio regarding Draw-Down/ Conditional Draw-Down At Risk.

Under the assumption of normally distributed returns, the wealth of a portfolio can be approximated by a Geometric Brownian Motion given by

$$X(t) = \sigma W(t) + \mu t$$

where W(t) is a standard Wiener process, μ is the drift and σ is the diffusion parameter. Now it is possible to derive the average Maximum Draw-Down. Its asymptotic behavior is

$$E[AD] = \frac{2\sigma^2}{\mu} Q_{AD}(\alpha^2)$$

$$Q_{AD}(x) \rightarrow \begin{cases} \mu < 0 & \begin{cases} x \rightarrow 0^+ & -\gamma\sqrt{2x} \\ x \rightarrow \infty & -x - \frac{1}{2} \end{cases} \\ \mu = 0 & 2\gamma\sigma\sqrt{T} \\ \mu > 0 & \begin{cases} x \rightarrow 0^+ & \gamma\sqrt{2x} \\ x \rightarrow \infty & \frac{1}{4} logx + 0.49088 \end{cases} \end{cases}$$

$$\alpha = \mu\sqrt(\frac{T}{2\sigma^2})$$
$$\gamma = \sqrt(\frac{\pi}{8})$$

with T as the investment horizon. This setup allows us to estimate the average Draw-Down of a time series by using its mean and variance. Again, we made the experience that in practise a lot of data is necessary to get a reasonable result. The reason might lie in the assumption of normality in the returns distribution which can be hold, if ever, only for very long time series.

If normality does not hold, the portfolio or assets wealth can not be modelled by a Geometric Brownian Motion and therefore the algebraic relation is not valid anymore.

7 Comparison Of The Risk Measures

Peijan and López de Prado state in [33] that there is also an algebraic relation between VaR, Draw-Down and Time Under-The-Water whenever normality and time-independence hold. In section 5.3 we have seen that variance is closely related to CVaR and in section 6.3 we have shown that there exists an algebraic correspondence between variance and Draw-Down for the case of normality.

This means that there is even an algebraic correspondence between Variance, Value at Risk, Draw-Down and Time Under-The-Water. And for the context of portfolio optimization we can conclude that the three optimization techniques minimizing the variance, minimizing CDaR and maximizing CVaR will end up with the same results if the assumptions of normality and time-independence hold.

This relation between the different risk measures disappears when normality can not be assumed anymore and we expect the optimization procedures to produce different results.

Part III
Optimization With Alternative Investments

This third part deals with the implementation of the discussed risk measures and the results achieved by using different data sets. Therefore we first show the implemented optimization problems and some numerical specialities related to them. Then we discuss quickly the different kind of data and show the used data. Afterwards the results of the calculations are shown and interpreted. Finally a summary and an outlook is given.

8 Numerical Implementation

The table below summarizes the considered optimization problems whereby μ indicates the expected return given by the investor.

$Min\ Variance$	$Max\ CVaR$	$Min\ CDaR$
$E[R] = \mu$	$E[R] = \mu$	$E[R] = \mu$
$\sum w_i = 1$	$\sum w_i = 1$	$\sum w_i = 1$
$w_i \geq 0$	$w_i \geq 0$	$w_i \geq 0$

In the following we list some aspects of the implementation:

- Since our optimizer is only capable of minimizing a function but not of maximizing a function, we minimize in the implementation $-CVaR$ instead of maximizing $CVaR$.
- We do not minimize the Variance but the standard deviation. Experiments have shown that this results in a better convergence of the solution. The reason might lie in the optimization algorithm that seeks the lowest value of a function by following the steepest gradient. Since we are mostly dealing with variances smaller than 1, the standard deviation - as square root of the variance - has a "broader minimum".
- In order to get the best results the efficient frontier gets calculated twice: A first run starts at the corner solution with the lowest expected return moving to the corner solution offering the highest expected return and afterwards a second run is executed in reverse order. The results of the first run are stored and compared with the results of the second run, whereby the better results (i.e. the portfolio weights leading to a smaller risk value) are chosen for the final output. While moving from one corner solution to the other, the optimal portfolios are calculated. The optimizer can be given an initial estimation for the weights of the optimal portfolio. These estimations of the optimal portfolio weights are calculated as linear extrapolation of the last two optimal portfolio weights because portfolio weights change often linearly while changing the expected return.

9 Used Data

In this section we will explain the difference between normal and logarithmic data and argue why we have decided to use logarithmic data. We also list the used historical market data and show how we have simulated artificial data from it.

9.1 Normal Vs. Logarithmic Data

In finance there are two common ways to model returns: Simple/geometric returns and logarithmic/ continuously compounded returns. The following table gives an overview of geometric and logarithmic returns for single- and multi-period each. $P_t and P_{t+1}$ denote the absolute value of the asset at time point t and t+1 respectively.

	Single-period	Multi-period
Geometric Return	$R_{t,t+1} = \frac{P_{t+1}}{P_t} - 1$	$R_{t,t+n} = [\prod_{i=0}^{n-1}(1 + R_{t+i,t+i+1})] - 1$
Logarithmic Return	$r_{t,t+1} = log(1 + R_{t,t+n})$	$r_{t,t+n} = log[\prod_{i=0}^{n-1}(1 + R_{t+i,t+i+1})]$ $= \sum_{i=0}^{n-1} r_{t+i,t+i+1}$

With geometric returns, the new return gets calculated at the end of each period and therefore the increase or decrease in the return gets active for the next period. In contrast, using continuously compounded returns, the change in the returns gets calculated on a infinitesimal small time period and therefore a continuously compounded return represents the actual value at every time point.

We have decided to use logarithmic/continuously compounded returns for the analysis for the following reasons:

- Because $0 \geq \frac{P_{t+n}}{P_t} \geq \infty$, using simple returns, the effective return can not be below -1 (full loss, for $P_{t+n} = 0$) which is an restriction to the range of possible values. Using logarithmic returns, the range gets stretched to $[-\infty, \infty]$. This is especially important for tail analysis, since the tail gets cut at -1 using simple returns and there would be a probability assigned to value that do not appear.

- If single-period returns are assumed to be normal, then multi-period returns $(\prod_i(1 + R_{t+i})) - 1$ are not normal. This comes from that the fact that a product of normally distributed variables is not normally distributed. By taking log-returns, multi-period returns are achieved by adding up the single-period returns which results again in a normal distribution (Central Limit Theorem)

- The concepts of CVaR, CDaR calculate thresholds that may lie in between a time period whereas the data is for the end of the time period. In this case it is more precise to use logarithmic returns instead of the linear approximation done by geometric returns.

9.2 Empirical Vs. Simulated Data

Empirical Market Data

As real market data we have chosen 3 bond indices, 5 equity indices and a hedge fund index. For equities and bonds there is a representative index for each of the following geographic categories: the whole world, Europe and the United States. Additionally we have also for equity indices for Far East and the Emerging Markets. As proxy for alternative investments the Hedge Fund Research (HFR) Fund Weighted Composite Index gets used. For a list of the various hedge fund styles included in this index and its descriptions you are referred to appendix F.
The data is coming from DataStream, except the HFR data coming directly from HFR. The data range includes almost the past 14 year (January 1990 until September 2003) on a monthly basis. This means that there are 165 data points per index available. The hedge fund index acts bottleneck because for all of the other indices more data into the past would be available. However to make the results more comparable, we restrict the data range to the largest common range. Not for all indices is it possible to get 10 years of data, e.g. the two indices based on the euro are just available after the introduction of this currency in 1999. Those indices not booked in USD, were converted to this currency. We are aware that, by converting all indices to USD, we have introduced currency risk to the time series. However, we think that it makes much more sense to compare time series that are in the same currency than different ones. In case that a value of a time series was missing for a certain date (e.g. because of a holiday), we have taken the value from the day before. The tests are applied to the log-returns of the data series.
The following table lists the indices and the first four moments of its logarithmic monthly returns.

Asset Class	Mean	Standard Deviation	Skewness	Excess Kurtosis
HFR Fund Weighted Composite	0.01140	0.0205	-0.775	3.24
MSCI World	0.00455	0.0435	-0.539	0.502
MSCI Europe	0.00603	0.0467	-0.566	0.847
MSCI North America	0.00866	0.0444	-0.569	0.600
MSCI Far East	-0.00374	0.0661	0.141	0.615
MSCI Emerging Markets	0.00533	0.0697	-1.08	3.19
JPM Global	0.00648	0.0191	0.505	0.972
JPM Europe	0.00681	0.0282	0.0245	0.744
JPM USA	0.00635	0.0131	-0.568	1.19

The Covariance matrix of the 9 asset classes is as follows:

	HFR FWC	MSCI WD	MSCI EU	MSCI US	MSCI FE	MSCI EM	JPM WD	JPM EU	JPM US
HFR FWC	0.000418	0.000604	0.000576	0.000636	0.000577	0.00108	-0.0000111	-0.0000705	-0.00000448
MSCI WD	0.000604	0.00188	0.00178	0.00161	0.00218	0.00205	0.000149	0.000120	0.00000066
MSCI EU	0.000576	0.00178	0.00216	0.00143	0.00167	0.00193	0.000213	0.000316	-0.00000169
MSCI US	0.000636	0.00161	0.00143	0.00196	0.00121	0.00195	0.0000218	-0.0000999	0.0000102
MSCI FE	0.000577	0.00218	0.00167	0.00121	0.00435	0.00227	0.000328	0.000301	0.0000152
MSCI EM	0.00108	0.00205	0.00193	0.00195	0.00227	0.00483	-0.0000991	-0.000262	-0.000156
JPM WD	-0.0000111	0.000149	0.000213	0.0000218	0.000328	-0.0000991	0.000362	0.000468	0.000164
JPM EU	-0.0000705	0.000120	0.000316	-0.0000999	0.000301	-0.000262	0.000468	0.000790	0.000170
JPM US	-0.00000448	0.00000066	-0.00000169	0.0000102	0.0000152	-0.000156	0.000164	0.000170	0.000172

Simulated data

We generate artificial data based on the historical data described above. For this purpose we first fit a multivariate skewed normal distribution and a multivariate skewed student-t distribution to the historical data. The fitting procedure gives us a vector of regression coefficients, the covariance matrix, a vector of shape parameters and the degree of freedom. In the case of fitting a skewed normal distribution the shape parameters are all 0 and the degree of freedom is infinite as it is well-know for the normal distribution.
Based on this estimated distributions we can generate random samples. As always with Monte Carlo Simulations, we have the advantage that we have full control over the underlying model because we can control and change the parameters. As disadvantage we note that the Monte Carlo ignores all dependencies over time in the time series and therefore slightly overstate the true value of diversification across assets classes in simulated portfolios.
Another unpleasant aspect is that there is only one value for the degree of freedom for all asset classes estimated and respected in the fitted function. This means that the time series do not have an individual kurtosis each but only a common one. However it is a non-trivial task to generate multivariate correlated data with skewness and kurtosis and would be beyond the scope of this thesis. One approach would be to use Copulas.

10 Evaluation Of The Portfolios

So far we have presented three methods how it could be possible to optimize a portfolio and in the last chapter we have introduced some historical market data. In the following part we will publish the portfolios that were optimized based on the market data. In a first section we show the results of the portfolio optimization if we use only traditional assets classes. Afterwards we introduce a hedge fund and analyze how it changes the optimal portfolios. In the third part we generate artificial data with the same characteristics as the asset classes and optimize this data.

For the portfolio optimization we give the expected return of the investor and try to minimize the respective risk. This procedure is done for several expected returns to get the efficient frontier. For the calculations, the range of these expected returns is defined as the interval between the smallest and the largest expected return of the assets classes. Clearly, under the assumption of no short sales and no lending and borrowing it is not possible to reach an expected portfolio return outside this interval (see chapter 1.2). We are aware that the part of the efficient frontier that is below the minimum risk portfolio is in practise not relevant. This is especially true for expected target returns below 0. We will nevertheless show the whole range to give the whole picture of the optimization results and to compare them.

For the charts we use the following color encoding:

Asset Class	Color	Style	Example
HFR Fund Weighted Composite	Black	Solid	
MSCI World	Orange	Solid	
MSCI Europe	Red	Solid	
MSCI North America	Green	Solid	
MSCI Far East	Blue	Solid	
MSCI Emerging Markets	Pink	Solid	
JPM Global	Orange	Dashed	
JPM Europe	Red	Dashed	
JPM USA	Green	Dashed	

The alpha value for the CVaR and CDaR optimization is chosen as 0.25, the size of the rolling window for the CDaR as 24 and the step size for the CDaR as 3. These are the values for which we got the most stable results. Since we have only 165 data points per time series, it was not possible to decrease the alpha value further towards 0.1 or use non-overlapping windows for the calculation of CDaR and still get reasonable results.

For the portfolio optimization no constraints for the weights were introduced in order to see the pure results and no influenced ones.

10.1 Evaluation With Historical Data

This section contains the results derived by using the original data series as presented before. All the calculations are done for the 8 traditional asset classes and again for the 9 asset classes including the hedge fund data.

Portfolios With Traditional Assets

Figure 30 shows the result of optimization of the 8 chosen traditional assets classes. The two pictures in the same row belong to the same optimization technique (Mean-Variance, Mean-CVaR or Mean-CDaR). The pictures in the left row show the weights of the individual assets classes dependent on the expected target return, chosen by the investor. The pictures in the right row show the efficient frontier resulting from the optimized weights.

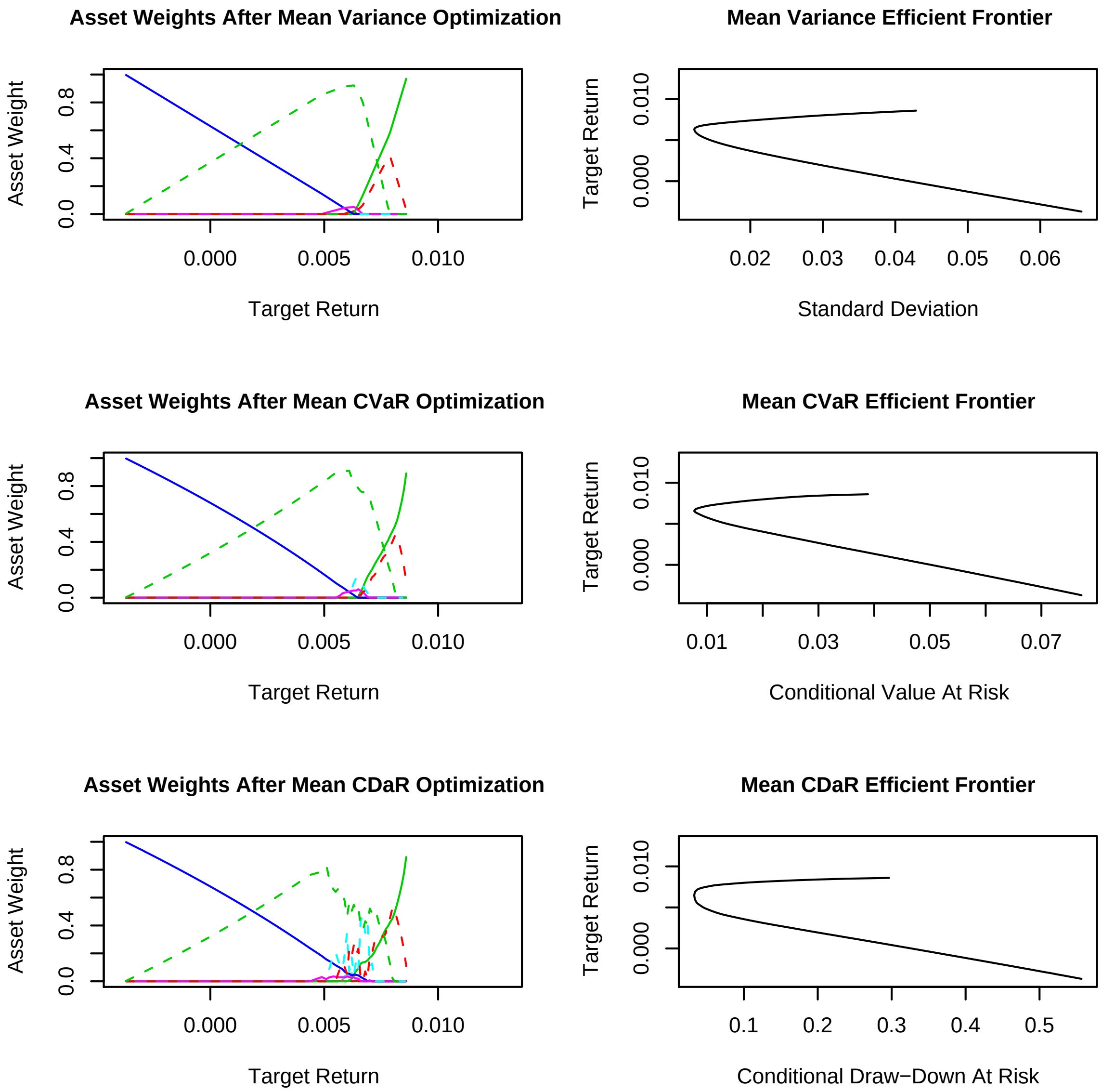


Figure 30: The weights and efficient frontiers for traditional asset classes for various optimization criteria.

At first sight we can see from the pictures that the optimization techniques produced very similar results. They all start with investing 100% of the available capital in MSCI Far East (blue solid line) if the investor asks for a very low return around -0.004. This is the only asset class the offers such a low mean return. As we increase the expected return the contribution from the JPM USA (green dashed line) increases until an expected return of 0.006 where the

contribution of MSCI Far East is decreased to a contribution of 0. In this area we can see differences in the asset allocation of the three techniques: The Mean Variance optimization pushes the JPM USA class to 100% to reach an expected return of 0.006, whereas Mean CDaR increases JPM USA to 0.8 at the maximum and distributes the resulting part to JPM global (turquoise doted line). All three techniques agree in the range above 0.007 to invest in JPM Europe (red doted line) and have a major allocation in MSCI North America when it comes to an expected return above 0.008.
The efficient frontier also look very similar for all three optimization techniques. The minimum risk portfolio is at an expected return of 0.0065 for all techniques. We can state that the efficient frontier of Mean Variance and Mean CVaR are more similar in comparison with the efficient frontier of Mean CDaR optimization.
The results do not correspond completely to the portfolio theory which says that in the area of lower expected return we can find mostly bonds because they offer usually a lower expected return and a low risk. In the higher region of expected returns we could expect equity indices from risky geographic locations as the Emerging Markets.
We can explain the calculated results with the actual situation at the world markets: The table with the four moments of the indices show that MSCI Far East is the only index with a negative first moment. The reason for this is the Asia crisis in 1997 that is contained in the data interval. The second moment shows us why indices like MSCI Emerging Markets and MSCI Europe don't appear in the weights chart: They have a too high Standard Deviation - especially in comparison to the bonds which offer a higher expected return for a lower Standard Deviation. Since the three optimizations are linked together via the standard deviation, this holds true for all of them.
The results also show the effect of diversification very clearly: MSCI Far East (blue solid line) and JPM US (green dashed line) which dominate the lower part of expected return have a correlation of -0.0001 (see Covariance matrix) and MSCI US (green solid line) together with JPM EU (red dashed line), which have a high allocation in the higher part of expected return, have a correlation of -0.0000152. These are two of the smallest entries in the Covariance matrix. This shows that all optimization techniques try to combine the fewest correlated assets.
This might also be the reason why MSCI World Equities is used so rarely to form the portfolios: MSCI World can be considered as a linear combination of the other indices. Since the optimization is looking for optimal diversification, the other indices, that inherit more extreme properties, are being used.
The results of the Mean-Variance optimization are very smooth, whereas the efficient frontier of the Mean-CDaR optimization is much more peaked and unstable. This effect might be coming from the small data set and the fact that CDaR (and also CVaR to a certain extent) take outliers heavily into consideration. As we will see, this artifacts will disappear as soon as we increase the amount of data.

Portfolios With Traditional And Alternative Assets

In this section we show the results of portfolio optimizations given that a hedge fund index is available. Figure 31 contains the six pictures with the weight allocation for the portfolios in the left row and the efficient frontier in the right row.

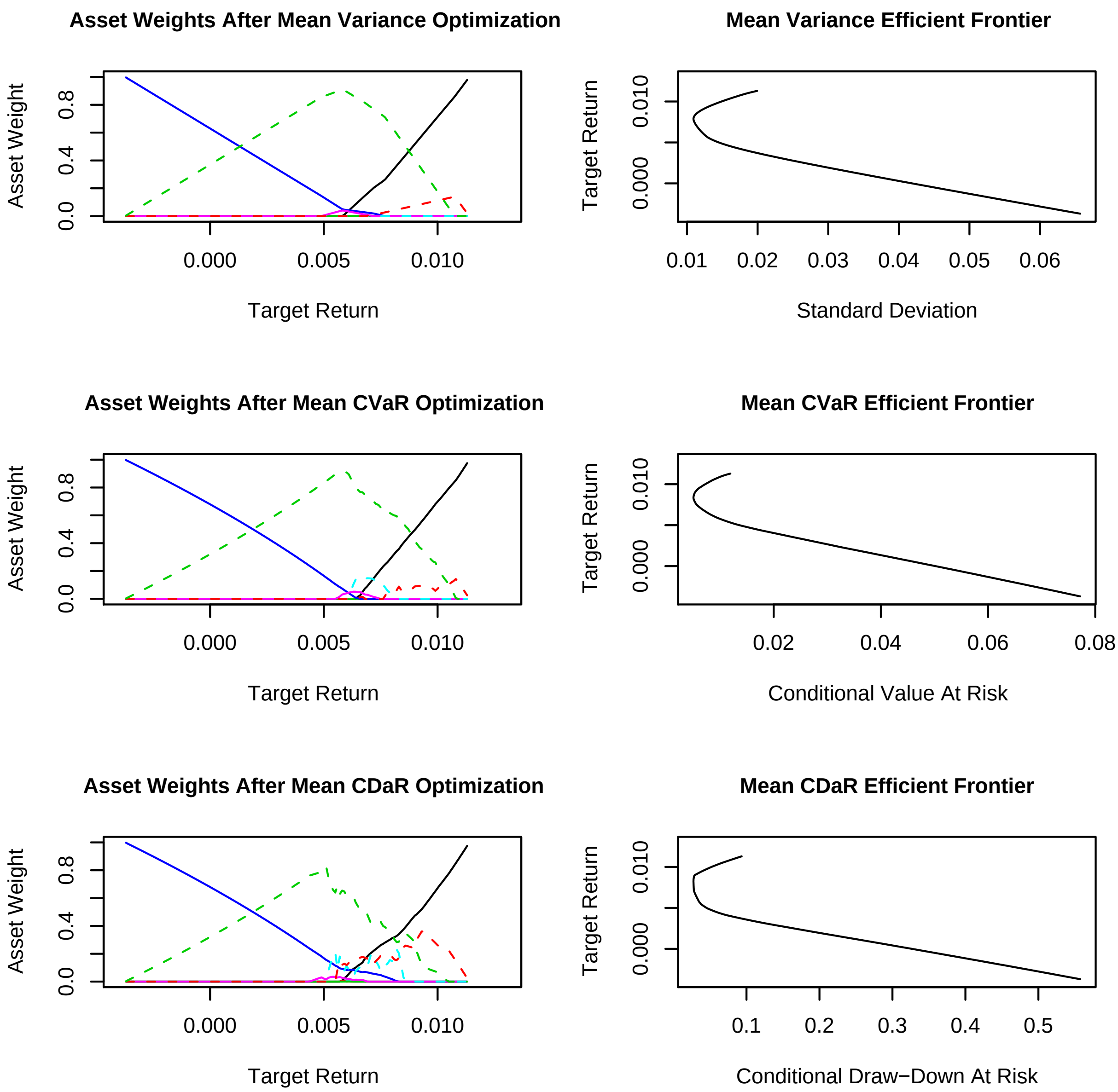


Figure 31: The weights and efficient frontiers for traditional and alternative asset classes for various optimization criteria.

Again it gets visible that the results of the optimization techniques are similar. Another interesting effect is that results for the expected returns in the range of -0.004 until 0.006 are the same for the case with and without the hedge fund index. This means that the hedge fund index has no influence to the lower expected returns but is treated as independent. As for the situation without hedge fund index, MSCI Far East (blue solid line) and JPM USA (green

dashed line) dominate the range between -0.004 and 0.006. Around the expected return of 0.005 we have for the Mean CDaR and Mean CVaR optimization also JPM Global (turquoise dashed line) playing a minor role. The hedge fund index gets taken into consideration when the expected return reaches a level of 0.006 and above. It attracts all the weight for expected returns above 0.010 because it is the only asset offering such a high return. Remarkable is that JPM Europe (red dashed line) gets over weighted in a Mean CDaR optimization in comparison to Mean Variance and Mean CVaR optimizations.
The Covariance matrix shows that the hedge fund index is very little correlated with the other assets. The results of the optimization suggest to combine the hedge fund index with JPM EU (red dashed line) and JPM US (green dashed line) to get a high expected portfolio return. The correlation of the hedge fund index is negative with both Bond indices. Again the effect of diversification got utilized by all of the optimization techniques.
The little peaks in the weight allocation charts show that the CDaR-results are much more instable compared to the Variance-Results.

10.2 Evaluation With Simulated Data

In this section we will optimize portfolios based on simulated data. As earlier described, the data is gained by fitting a distribution to the available monthly time series of the assets classes. We distinguish between fitting a skewed normal distribution and fitting a skewed student-t distribution. As soon as we have the distribution, we can generate as many artificial data with the same properties as we need. For the following calculations we have generated 2000 samples for each asset class. This represents 2000 months or 167 years of data.

Portfolios With Simulated Traditional Assets

Figure 32 shows that the results we get when fitting a multivariate skew normal distribution to the historical data and generating 2000 samples with this distribution are pretty much similar to the ones of the original data. We can see that the instability in the CDaR and CVaR optimization disappears and all of the three optimizations get the same results. Only a little peak of 10 percent allocation in MSCI Emerging Markets in the CDaR optimization distinguishes the results.
In Figure 33 the results for fitting a multivariate student-t distribution to the same 8 data series are provided. The covered range for the expected return has shifted from the interval (-0.004, 0.008) to the interval (-0.001, 0.010) which is a results of the randomly generation of new data from the fitted distribution. Besides this shift there is another difference compared to fitting a skewed normal distribution: The allocation of JPM Europe (dotted red line) is more varying comparing the three optimization techniques. It fluctuates from 20 percent for Mean-Variance optimization up to almost 40 percent for Mean-CDaR optimization. This effect might be coming from the fitted skewed student-t distribution that allows a higher adaptation to the original data then the skewed normal distribution.

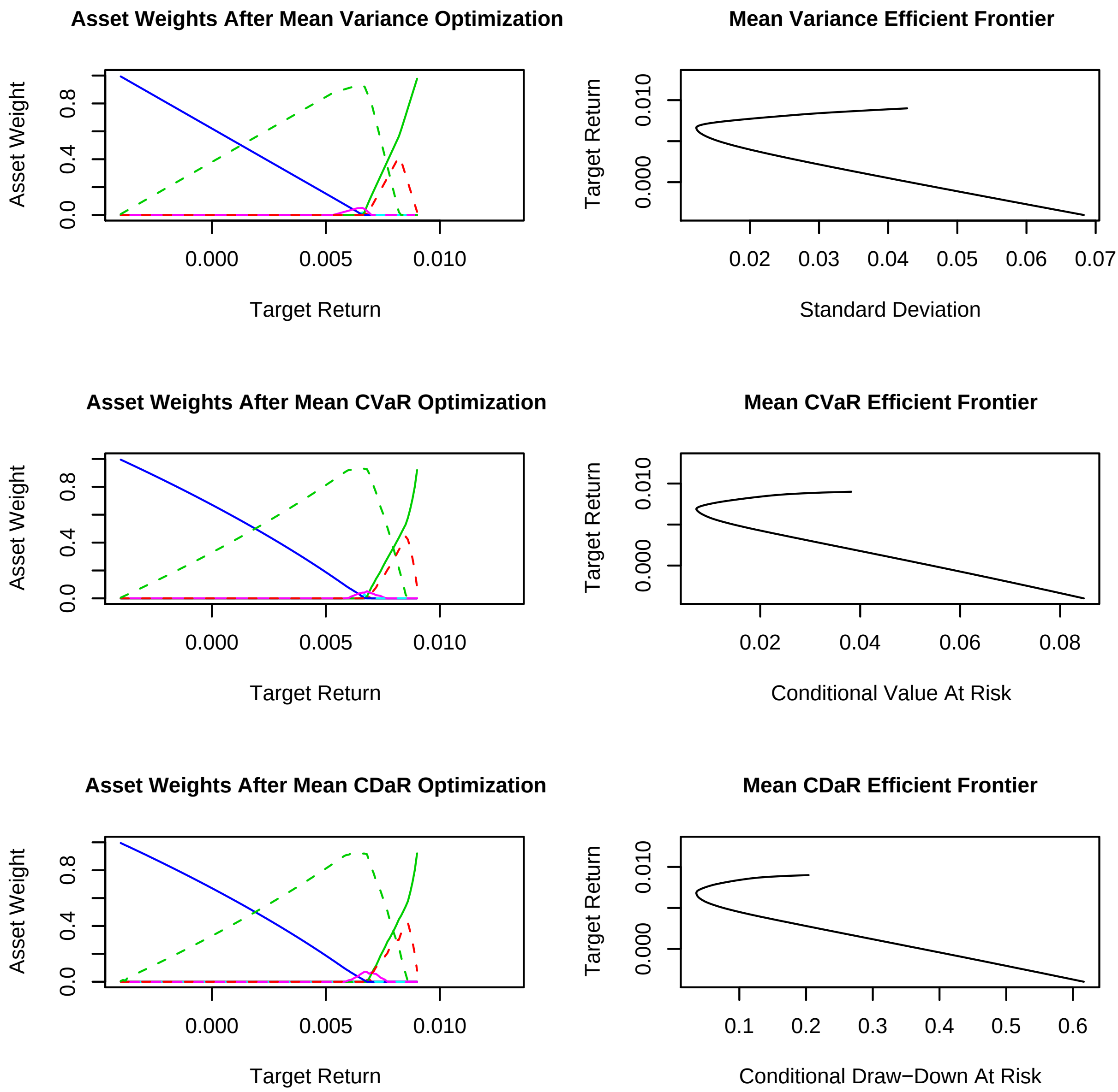


Figure 32: The weights and efficient frontiers for traditional asset classes for various optimization criteria. The used data has been simulated by a skewed normal distribution fitted to the historical data.

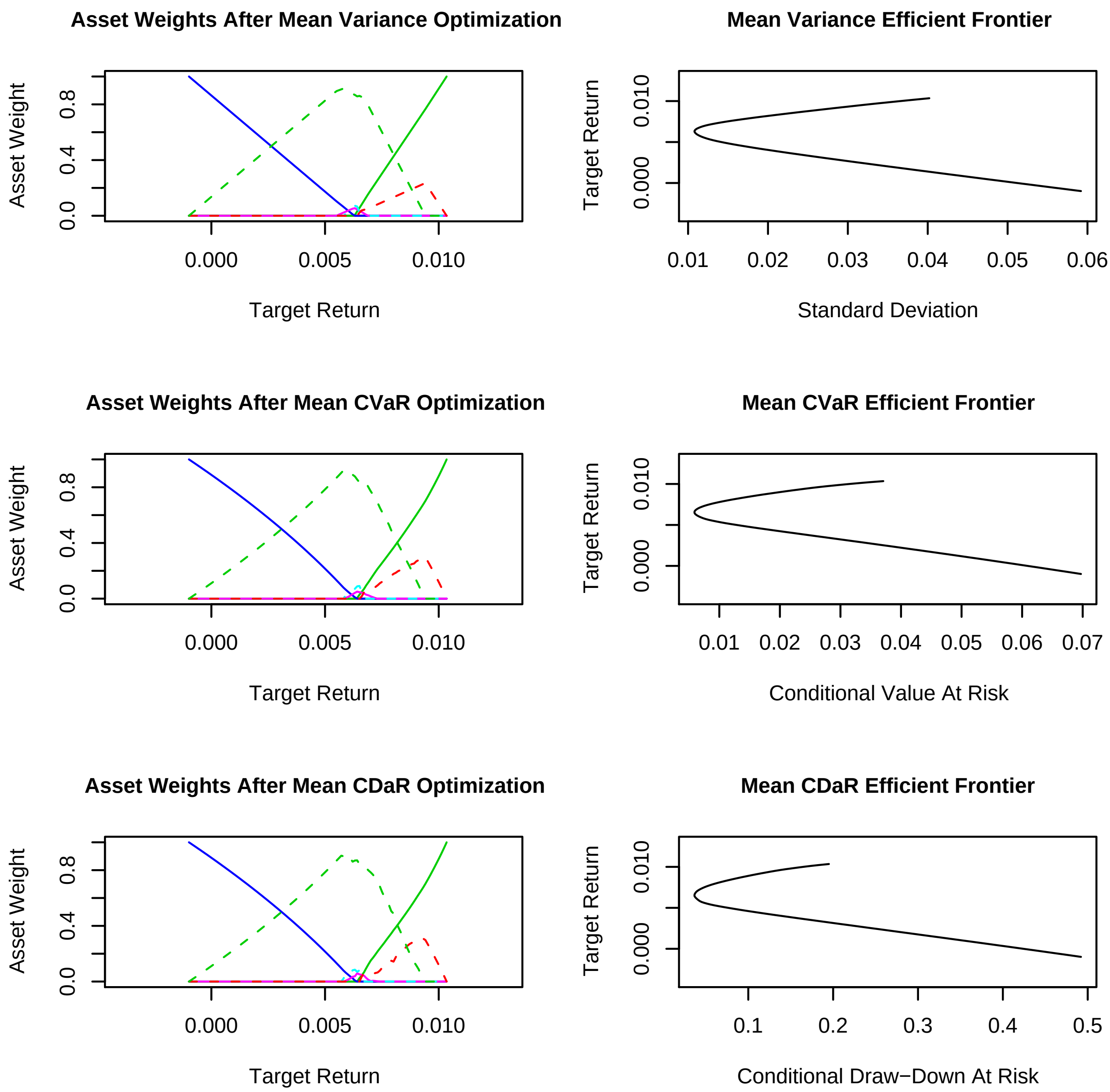


Figure 33: The weights and efficient frontiers for traditional asset classes for various optimization criteria. The used data has been simulated by a skewed student-t distribution fitted to the historical data.

Portfolios With Simulated Traditional Assets And Alternative Assets

Figure 34 and figure 35 show the result for the 9 assets classes, including the hedge fund index. The results of figure 34 are retrieved by fitting a skewed normal distribution to the historical data, whereas the results of figure 35 are retrieved by fitting a skewed student-t distribution to the historical data.

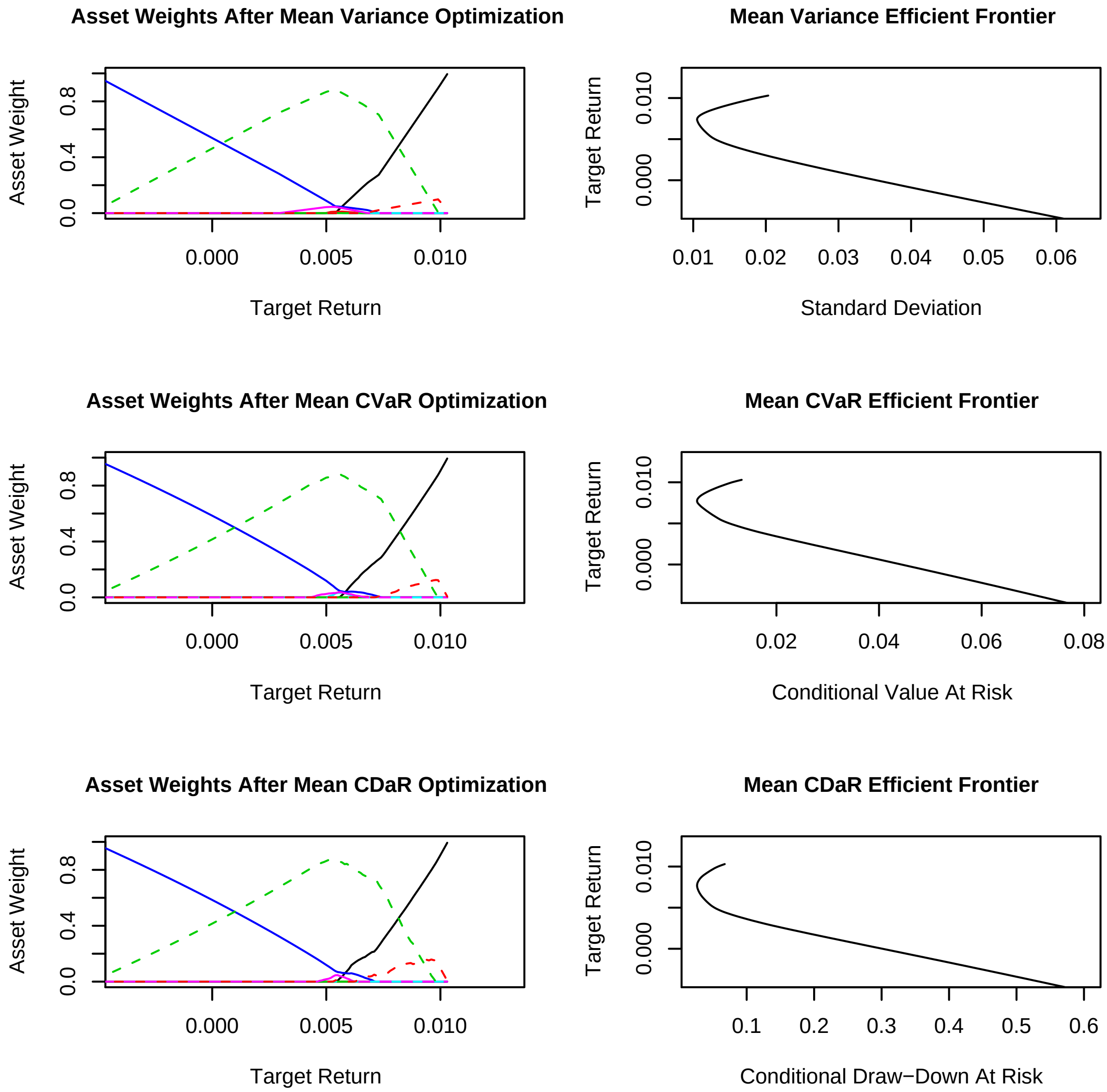


Figure 34: The weights and efficient frontiers for traditional and alternative asset classes for various optimization criteria. The used data has been simulated by a skewed normal distribution fitted to the historical data.

Comparing figure 34 and figure 35 we see again the same effect as we have seen for the 8 asset classes: the outcome of the three different optimizations differs more when we fit the historical data with a multivariate skewed student-t distribution instead of the multivariate skewed normal distribution. Besides this we can confirm that hedge funds offer a possibility for higher returns.

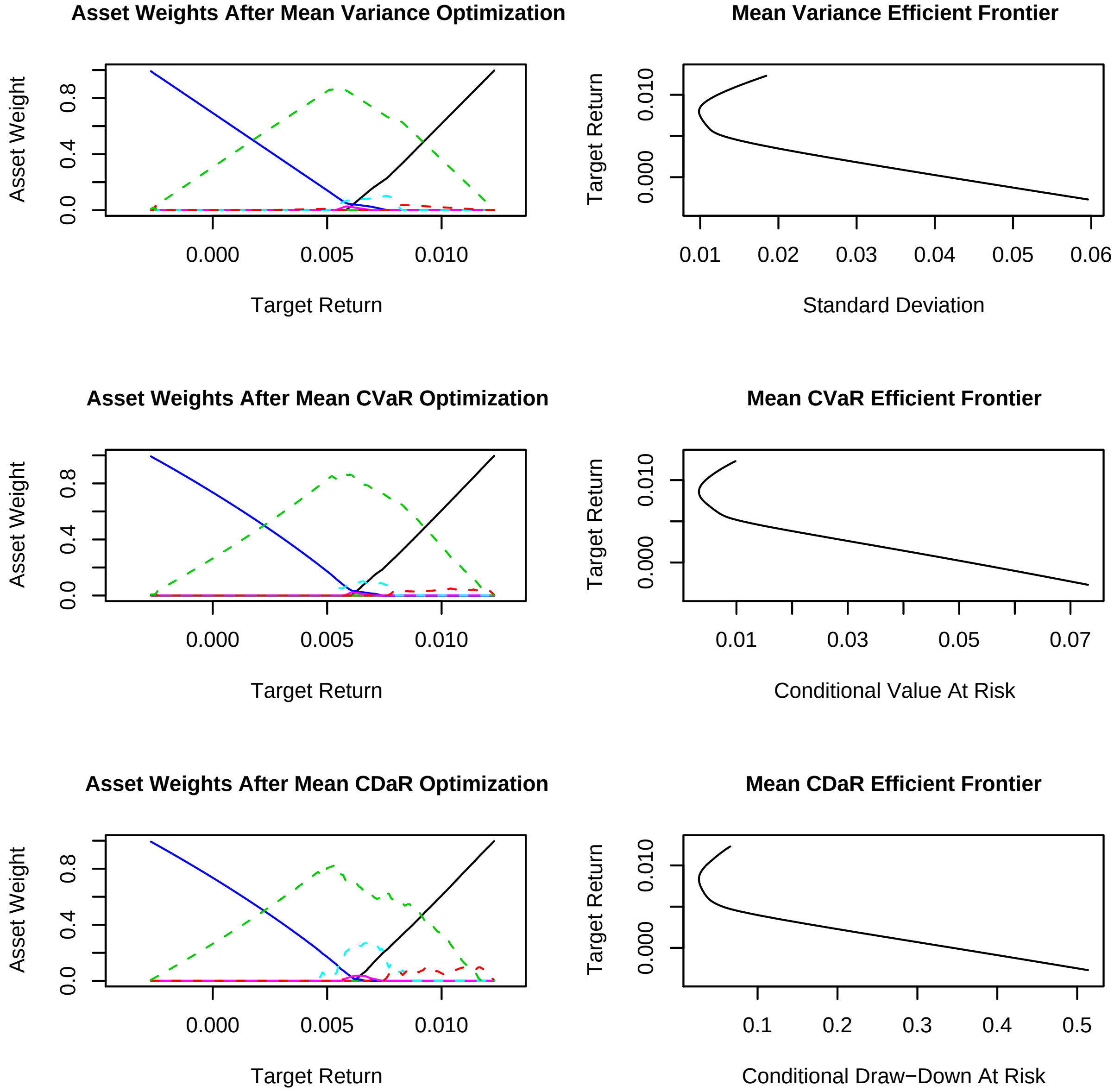


Figure 35: The weights and efficient frontiers for traditional and alternative asset classes for various optimization criteria. The used data has been simulated by a skewed student-t distribution fitted to the historical data.

Summary and Outlook

Portfolio optimization had always been a key issue of finance. In recent years its complexity increased because of the emergence of derivatives and alternative instruments. New alternative investment vehicles like hedge funds are very interesting in the context of portfolio optimization because they offer a lot of unexplored investment opportunities. This thesis dealt with the question of how to integrate alternative investments like hedge funds into a portfolio.

In the first part we presented the standard portfolio optimization approach according to Markowitz by describing the risk return framework and the relation to the utility function of an investor. Important here is to state that the standard Mean-Variance optimization assumes normal distributed returns or a specific utility function for the investor. The analytical solutions for optimal portfolios were derived for the case of two assets.

The purpose of the second part was to show that the requirements of the Mean-Variance optimization as proposed by Markowitz are not completely fulfilled and to present some alternative optimization processes. To show the violation of the requirements, we applied some statistical tests for measuring the stylized facts of asset returns. The numerical results showed that the returns are not normal distributed but have fat tails. The stylized facts appear especially strong when we increase the data frequency (e.g. going from monthly data to daily data). Afterwards we discussed the pleasant properties of risk measures and present several sets of properties as proposed in literature. In order to propose alternatives to the Mean-Variance optimization, Value at Risk, Draw-Down and Time Under-The-Water and its derivations Conditional Value at Risk and Conditional Draw-Down at Risk were introduced. They were analyzed and compared with the variance as risk measure. It is explained that portfolio optimized according to variance, Value at Risk or Draw-Down will be very similar in the case of normal distributed data.

The third part summarized the results achieved by applying the three optimization techniques Mean-Variance, Mean-Conditional Value at Risk, and Mean-Conditional Draw-Down at Risk to data. For this purpose we have implemented a software framework to test and compare the different optimization techniques. This software framework and also the used data is explained. We have introduced historical hedge fund data because is known that hedge fund returns exhibit special statistical properties like skewness and kurtosis and it is therefore interesting to see how they influence the portfolio optimization results. The data were twofold: We used empirical data and simulated time series based on fitting multivariate skewed distribution functions to the empirical returns. For each setup of data and optimization technique we have calculated the efficient frontier and the weight allocation of the efficient portfolios. The results of the three optimization techniques differed dependent on the used data. As expected was the outcome for the different optimization techniques less variable for the case of normal data and was more varying when we used non-normal data. This supported the conclusion from the algebraic analysis of the risk measures that portfolio optimization techniques different than the Mean-Value optimization are preferable in the context of non-normal data. Therefore we propose to use risk measures like Conditional Value at Risk or Conditional Draw-Down at Risk especially in the case of alternative investments because their returns deviate from the normal distribution. However hedge funds have also good properties if one wants to go on with the Mean-Variance optimization: In the investigated period hedge funds had a very good performance and offer therefore a very high return. Even if the performance will decrease in the future (e.g. because of stricter regulations), hedge funds will still be a very good way to diversify a portfolio because of the low correlation with the traditional assets.

The implemented software features efficient algorithms and interfaces to other programming languages. It is modularly designed in order to get the code easily changed and the functionality enhanced. We think that risk measures can be comfortable discovered and analyzed with this software. It would be interesting to use other kind of data and implement new risk measures.

Appendix

A Quadratic Utility Function Implies That Mean Variance Analysis Is Optimal

In this appendix we want to show that it is possible to express the expected utility function in terms of mean and variance and that it is therefore optimal to apply an mean variance analysis if one uses a quadratic utility function.

The variance of a random variable W is in (2) defined as

$$\sigma_W^2 = E[W - E[W]^2] = E[W^2 - 2W * E[W] + E[W]^2]$$

Because

$$E[\sum_{i=1}^{N} X_i] = \sum_{i=1}^{N} E[X_i]$$

holds, we get

$$\sigma_W^2 = E[W^2] - E[2W * E[W]] + E[W]^2$$

and since

$$E[c * X] = c * E[X]$$

holds, we can rewrite the variance as

$$\sigma_W^2 = E[W^2] - 2 * E[W] * E[W] + E[W]^2 = E[W^2] - [E[W]]^2$$

Rearranging yields to

$$E[W^2] = \sigma_W^2 + [E[W]]^2$$

We have the expected value of the quadratic utility function we want to optimize

$$E[U(W)] = E[W] - b * E[W^2]$$

Here we can substitute the term derived two lines above and get

$$E[U(W)] = E[W] - b * [\sigma_W^2 + [E[W]]^2]$$

Deriving this term we have proven that, assuming a quadratic utility function, a mean variance analysis optimizes the expected utility.

B Equivalence Of Different VaR Definitions And Notations

Definitions and Notations used in this thesis:

$$VaR_\alpha = sup\{x|P[R_P < x] \leq \alpha\} \tag{35}$$

where α is expected to be in [0.01, 0.1] and x is a random variable of a return function.

$$CVaR_\alpha = E[R_P|R_P \leq VaR] \tag{36}$$

This notation corresponds to the left graphic of figure 36.
In contrast we find in [35] and [36] the following definitions and notations

$$VaR_{1-\alpha} = inf\{x|P[R_P \leq x] \geq \alpha\} \tag{37}$$

where α is expected to be in [0.9, 0.99] and x is a random variable of a loss function.

$$CVaR_{1-\alpha} = E[R_P|R_P \leq VaR] \tag{38}$$

This corresponds to the right graphic of figure 36.

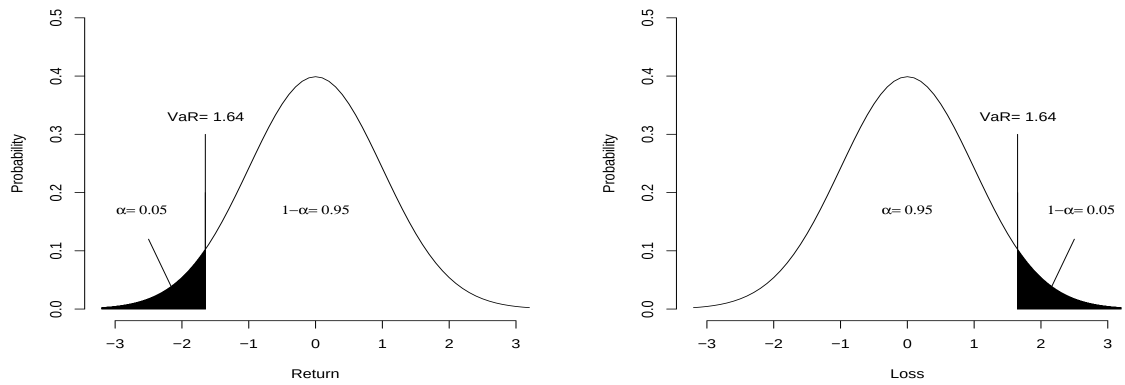


Figure 36: The two graphics depict the situation for the two kind of definitions of VaR and CVaR for the case of a standard normal distribution and a 5%/95% confidence level. The left graphic shows an $\alpha = 0.05$ and a return function. The right graphic depicts an $\alpha = 0.95$ and a loss function

The formulas (35) and (36) are defined on return functions (a positive value means a high return, a negative value indicates a loss) and calculates the 5% quantile whereas formulas (37) and (38) are defined on loss functions (a positive value means a loss, a negative value indicates a gain) and deals with the 95% quantile

The transformations of the VaR and CVaR can be expressed as

$$VaR_\alpha(X) = VaR_{(1-\alpha)}(-X) \tag{39}$$

$$CVaR_\alpha(X) = CVaR_{(1-\alpha)}(-X) \tag{40}$$

In [1], [2] [3] we can find a mixture of both notations where the same definitions as formulas (35) and (36) are used with a negative sign for both formulas in order to comply with the sign of formulas (37) and (38).

C Used R Functions

The following functions of the R programming language and environment were used for the implementation of the software system:

Function	Package	Description
apply	base	Returns a vector or array or list of values obtained by applying a function to margins of an array
arima.sim	ts	Simulate from an ARIMA model
bds.test	tseries	Computes and prints the BDS test statistic for the null that 'x' is a series of i.i.d. random variables
data.csv	fBasics	Loads specified data sets, or lists the available data sets
floor	base	Rounding of Numbers
garchSim	fSeries	Univariate GARCH time series modelling
length	base	Get or set the length of vectors (including lists)
lines	base	Add Connected Line Segments to a Plot
ksgofTest	fBasic	Performs a Kolmogorov-Smirnov Goodness-of-Test
mean	base	Generic function for the (trimmed) arithmetic mean
msn.fit	sn	Fits a multivariate skew-normal (MSN) distribution to data
mst.fit	sn	Fits a multivariate skew-student-t (MST) distribution to data
plot	base	Generic function for plotting of R objects
qnorm	base	Quantile function generation for the normal distribution
qqPlot	fExtremes	Produces a Quantile-Quantile plot of two data sets
qt	base	Quantile function generation for the t distribution
rmsn	sn	Random number generation for the multivariate skew-normal distribution
rmst	sn	Random number generation for the multivariate skew-student distribution
rmvnorm	mvtnorm	Generates random deviates from the multivariate normal distribution
rmvt	mvtnorm	Generates random deviates from the multivariate student distribution
rnorm	base	Random generation for the normal distribution
rsn	sn	Random number generation for the skew-normal distribution
rst	sn	Random number generation for the skew-student-t distribution
rt	base	Quantile function generation for the t distribution
runif	base	Generates random deviates from the uniform distribution
runsTest	fBasics	Performs a Runs Test
sum	base	Returns the sum of all the values present in its arguments
var	base	Computes the variance

D Description Of The Portfolio Optimization System

It was our intension to do all the calculations on a common hard-/software system in order to make the analysis as useful for practical applications as possible and easy for future extensions. This justifies the following system:

- The system runs on current personal computers (3GHz clock cycles, 1GB memory). We don't assume the availability of a supercomputer or pc-cluster.
- As software components we use R as front-end application and for some small calculations and an optimizer module written in Fortran77.

We will now describe how we have designed the system for portfolio optimization. The optimizer is written in Fortran77 which can be executed directly from R. The full system works as follows (see figure 37): R calls the optimization routine DONLP2 and gives the needed data (id of optimization method, asset returns, expected return of portfolio) as parameter to the optimizer. The optimizer itself calls several subroutines that define the objective function, equality constraints and inequality constraints and all of its gradients. In case that it is not possible to define analytic gradient functions, we have implemented a numerical gradient function.

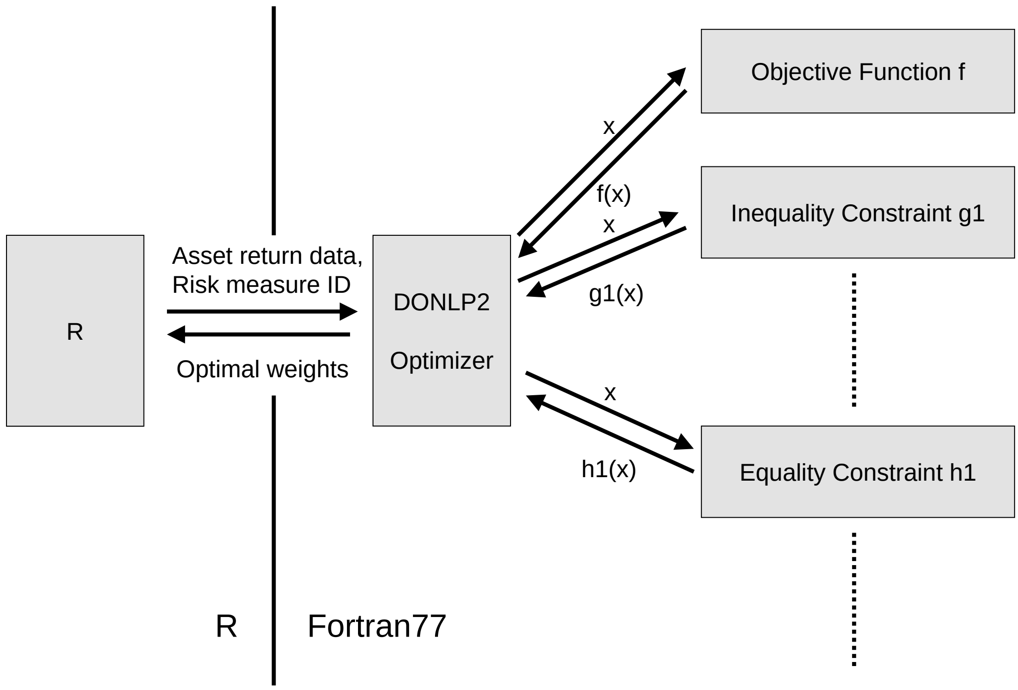


Figure 37: Schema of the dependencies of the optimization process.

Our intension is to develop a general purpose system that can easily be installed and extended. For this reason we have chosen a general non-linear optimizer that can be applied to any kind of problems. We are aware that it could be more time efficient to use specialized optimizers for each problem (e.g. a linear optimizer for the Conditional Value at Risk problem), however we think that the overhead of a general optimizer is negligible in our context.

The used optimizer 'DONLP2' can be downloaded for free from http://ftp.mathematik.tu-darmstadt.de/pub/department/software/opti/ where it is available as Fortran or C implementation. The correct functionality of the optimizer was tested with cross-tests to the optimizer in the R-package "quadprog" and the optimizer included in Microsoft Excel. In the documentation DONLP2 is described as

Purpose:
Minimization of an (in general nonlinear) differentiable real function f subject to (in general nonlinear) inequality and equality constraints g, h.

$$f(x) = min_{x \in S}$$

$$S = \{x \in R^n : h(x) = 0, g(x) \geq 0\}$$

Here g and h are vectorvalued functions.
Bound constraints are integrated in the inequality constraints g. These might be identified by a special indicator in order to simplify calculation of its gradients and also in order to allow a special treatment, known as the gradient projection technique. Also fixed variables might be introduced via h in the same manner.

Method employed:
The method implemented is a sequential equality constrained quadratic programming method (with an active set technique) with an alternative usage of a fully regularized mixed constrained subproblem in case of nonregular constraints (i.e. linear dependent gradients in the "working set"). It uses a slightly modified version of the Pantoja-Mayne update for the Hessian of the Lagrangian, variable dual scaling and an improved Armijo-type stepsize algorithm. Bounds on the variables are treated in a gradient-projection like fashion. Details can be found in [40] and [41].

E Description Of The Excel Optimizer

Optimization in Microsoft Excel begins with an ordinary spreadsheet model. The spreadsheets formula language functions as the algebraic language used to define the model. Through the Solvers GUI, the user specifies an objective and constraints by pointing and clicking with a mouse and filling in dialog boxes. The Solver then analyzes the complete optimization model and produces the matrix form required by the optimizers. The optimizers employ the simplex, generalized-reduced-gradient, and branch-and-bound methods to find an optimal solution and sensitivity information. The solver uses the solution values to update the model spreadsheet and provides sensitivity and other summary information on additional report spreadsheets.

Detailed information about the methods applied in the optimizer included in Microsoft Excel are given by Fylstra et al [22].

F Description Of Various Hedge Fund Styles

This section lists and explains some common hedge fund strategies. The strategies are taken from [6] and the respective volatility classification from the webpage www.magnum.com.

- **Convertible Arbitrage.** Expected Volatility: Low
  Attempts to exploit anomalies in prices of corporate securities that are convertible into common stocks (convertible bonds, warrants and convertible preferred stocks). Convertible bonds tends to be under-priced because of market segmentation; investors discount securities that are likely to change types: if the issuer does well, the convertible bond behaves like a stock; if the issuer does poorly, the convertible bond behaves like distressed debt. Managers typically buy (or sometimes sell) these securities and then hedge part or all of the associated risks by shorting the stock. Delta neutrality is often targeted. Overhedging is appropriate when there is concern about default as the excess short position may partially hedge against a reduction in credit quality.

- **Dedicated Short Bias.** Expected Volatility: Very High
  Sells securities short in anticipation of being able to re-buy them at a future date at a lower price due to the managers assessment of the overvaluation of the securities, or the market, or in anticipation of earnings disappointments often due to accounting irregularities, new competition, change of management, etc. Often used as a hedge to offset long-only portfolios and by those who feel the market is approaching a bearish cycle.

- **Emerging Markets.** Expected Volatility: Very High
  Invests in equity or debt of emerging (less mature) markets that tend to have higher inflation and volatile growth. Short selling is not permitted in many emerging markets, and, therefore, effective hedging is often not available, although Brady debt can be partially hedged via U.S. Treasury futures and currency markets.

- **Long/Short Equity.** Expected Volatility: Low
  Invests both in long and short equity portfolios generally in the same sectors of the market. Market risk is greatly reduced, but effective stock analysis and stock picking is essential to obtaining meaningful results. Leverage may be used to enhance returns. Usually low or no correlation to the market. Sometimes uses market index futures to hedge out systematic (market) risk. Relative benchmark index is usually T-bills.

- **Equity Market Neutral.** Expected Volatility: Low
  Hedge strategies that take long and short positions in such a way that the impact of the overall market is minimized. Market neutral can imply dollar neutral, beta neutral or both.

  - Dollar neutral strategy has zero net investment (i.e., equal dollar amounts in long and short positions).
  - Beta neutral strategy targets a zero total portfolio beta (i.e., the beta of the long side equals the beta of the short side). While dollar neutrality has the virtue of simplicity, beta neutrality better defines a strategy uncorrelated with the market return.

  Many practitioners of market-neutral long/short equity trading balance their longs and shorts in the same sector or industry. By being sector neutral, they avoid the risk of market swings affecting some industries or sectors differently than others.

- **Event Driven.** Expected Volatility: Moderate
  Corporate transactions and special situations
  - Deal Arbitrage (long/short equity securities of companies involved in corporate transactions)
  - Bankruptcy/Distressed (long undervalued securities of companies usually in financial distress)
  - Multi-strategy (deals in both deal arbitrage and bankruptcy)
- **Fixed Income Arbitrage.** Expected Volatility: Low
  Attempts to hedge out most interest rate risk by taking offsetting positions. May also use futures to hedge out interest rate risk.
- **Global Macro.** Expected Volatility: Very High
  Aims to profit from changes in global economies, typically brought about by shifts in government policy that impact interest rates, in turn affecting currency, stock, and bond markets. Participates in all major markets equities, bonds, currencies and commodities though not always at the same time. Uses leverage and derivatives to accentuate the impact of market moves. Utilizes hedging, but the leveraged directional investments tend to have the largest impact on performance.
- **Managed Futures.**
  Opportunistically long and short multiple financial and/or non financial assets. Sub-indexes include Systematic (long or short markets based on trend-following or other quantitative analysis) and Discretionary (long or short markets based on qualitative/fundamental analysis often with technical input).

G References

References

[1] Acerbi C., Nordio C., Sirtori C., 2001: *Expected Shortfall as a Tool for Financial Risk Management.*
www.gloriamundi.com

[2] Acerbi C., Tasche D., 2001: *Expected Shortfall: a natural coherent alternative to value at risk*
www.gloriamundi.com

[3] Acerbi C., Tasche D., 2001: *On the Coherence of Expected Shortfall.*
www.gloriamundi.com

[4] Albrecht P., 2003: *Risk Measures*
Contribution prepared for: Encyclopedia of Actuarial Science
John Wiley & Sons

[5] Alexander C., 2001: *Market models*
John Wiley &Sons

[6] Amenc N., Martellini L., 2002: *The Brace New World of Hedge Fund Indexes*
Working Paper

[7] Artzner P., Delbaen F., Eber J.-M., Heat D., 1999: *Coherent measures of risk*
Mathematical Finance 9, pp. 203-228

[8] Belaire-Franch, J., Contrras-Bayarri D., 2002: *The BDS Test: A Practioner's Guide*
Journal of Applied Econometrics 17, pp. 691-699

[9] Bernoulli D., 1738: *Exposition of a new theory on the measurement of risk.*

[10] Bollerslev T., 1986: *Generalized Autoregressive Conditional Heteroscedasticity.*
Journal of Econometrics

[11] Brock W., Dechert W.D., Scheinkman J., 1987: *A Test for Independence Based on the Correlation Diemension.*
University of Wisconsin Working Paper No. 8702

[12] Cheklov A., Uryasev S., Zabarankin M., 2003: *Portfolio Optimization With Drawdown Constraints.*
Working Paper

[13] Cheridito P, Delbaen F., Kupper M., 2003: *Coherent and convex risk measures for bounded càdlàg˙*

[14] Conover W.J., 1999: *Practical Nonparametric Statistics.*
John Wiley & Sons, Inc

[15] De Giorgi E., 2002: *A Note on Portfolio Selctions under Various Risk Measures.*
Working Paper Series ISSN 1424-0459

[16] Dembo R., King A., 1992: *Tracking Models and the Optimal Regret Distribution in Asset Location.*
Applied Stochastic Models and Data Analysis 8, pp. 151-157

[17] Dembo R., Freeman A., 2001: *The Rules of Risk.*
John Wiley & Sons, Inc

[18] Elton E., Gruber M., 1995: *Modern portfolio theory.*
John Wiley & Sons, Inc

[19] Embrechts P., Klueppelberg C., Mikosch T., 2002: *Modelling Extremal Events.*
Springer, pp. 290-294

[20] Embrechts P., Höing A., Juri A., 2003: *Using Copulae to bound the Value-at-Risk for functions of dependent risks.*
Finance & Stochastics 7, pp.145-167

[21] Fishburn P., 1977: *Mena-risk analysis with risk associated with below-target returns.*
The Amerian Economic Review 67, pp. 116-126

[22] Fylstra D., Lasdon L., Watson J., Waren A., 1998: *Design and Use of the Microsoft Excel Solver*
Interfaces 28, pp.29-55

[23] Gaivoronski A., Pflug G., 2000: *Value at Risk in Portfolio Optimization*

[24] Grinold R., Kahn R., 1999: *Active portfolio managemet*
McGraw-Hill, p. 99

[25] Huerlimann W., 2001: *Conditional Value-at-Risk bounds for Compund Poisson Risks and Normal Appriximation.*
Working Paper, MPS: Applied mathematics/0201009

[26] Jia J., Dyer S., 1996: *A Standard Measure of Risk and Risk-Value Models*
Management Science 42, pp. 1691-1705

[27] Kritzman M., 1995: *The portable financial analyst.*
The Financial Analysts Journal

[28] Magdon-Ismail M., Atiya A., Pratap A., Abu-Mostafa Y., 2003: *On the Maximum Drawdown of a Brownian Motion*
Working Paper

[29] Mandelbrot B., 1963: *The variation of certain speculative prices.*
Journal of Business 36, pp. 394-419

[30] Markowitz H., 1959: *Portfolio selection: Efficient diversification of investments.*
John Wiley & Sons

[31] de Moivre A., 1733:

[32] Pedersen C.S., Satchell E., 1998: *An extended family of financial risk measures*
Geneva Papers on Risk and Insurance Theory 23, pp. 89-117

[33] Peijan A., López de Prado M., 2003: *Measuring Loss Potential Of Hedge Fund Strategies.* Working Paper UBS Wealth Management and Business Banking

[34] Pflug G., 2000: *Some Remarks on the value-at-risk and the conditional value-at-risk.* Working Paper at Department of Statistics and Decision Support Systems, University of Vienna

[35] Rockafellar R., Uryasev S., 1999: *Optimization of Conditional value-at-risk* www.gloriamundi.com

[36] Rockafellar R., Uryasev S., 2002: *Conditional value-at-risk for general loss distributions* Journal of Banking & Finance 26 www.gloriamundi.com

[37] Ross S., 1976: *The arbitrage theory of capital asset pricing.* Journal of Economic Theory, 13, pp. 341-360

[38] Roy A., 1952: *Safety Firts and the Holding of Assets.* Econometrica 20, pp.431-449

[39] Schneeweiss H., 1967: *Entscheidungskriterien bei Risiko.* Springer, pp. 113-117

[40] Spellucci P., 1998: *An SQP method for general nonlinear programs using only equality constrained subproblems* Math. Prog. 82, (1998), pp. 413-448 Physica Verlag, Heidelberg, Germany

[41] Spellucci P., 1998: *A new technique for inconsistent problems in the SQP method* Math. Meth. of Oper. Res. 47, pp. 355-400 Physica Verlag, Heidelberg, Germany

[42] Stone B., 1973: *A General Class of Three-Parameter Risk Measures.* Journal of Finance 28, pp.675-685

[43] Testuri C., Uryasev S., 2002: *On Relation Between Expected Regret and Condtional Value-at-Risk.* Working Paper, University of Florida www.gloriamundi.com

[44] Uryasev S., 2000: *Conditional Value-at-Risk (CVaR): Algorithms and Applications.* Working Paper, University of Florida www.ise.ufl.edu/uryasev

[45] Wang S., Young V., Panjier H., 1997: *Axiomatic characterization of insurance prices.* Insurance: Mathematics and Economics 21, pp. 173 - 183