Does copula beat linearity?

Comparison of copulas and linear correlation in portfolio optimization.

Joakim Blom
Joakim Wargclou
DOES COPULA BEAT LINEARITY?
Comparison of copulas and linear correlation in portfolio optimization.

Blom, Joakim & Wargelou, Joakim.

Submitted in partial fulfillment of the requirements for the degree
Master of Science in Industrial Engineering and Management Department
of Mathematics and Mathematical Statistics.
Umeå University
SE-901 87 Umeå, Sweden.

Supervisor: Lisa Hed, Umeå University.

Examiner: Niklas Lundström, Umeå University.
Abstract

Modern portfolio theory (MPT) is an investment theory which was introduced by Harry Markowitz in 1952 and describes how risk averse investors can optimize their portfolios. The objective of MPT is to assemble a portfolio by maximizing the expected return given a level of market risk or minimizing the market risk given an expected return. Although MPT has gained popularity over the years it has also been criticized for several theoretical and empirical shortcomings such as using variance as a measure of risk, measuring the dependence with linear correlation and assuming that returns are normally distributed when in fact empirical data suggests otherwise. When moving away from the assumption that returns are elliptical distributed, for example normally distributed, we can not use linear correlation as a measure of dependence in an accurate way. Copulas are a flexible tool for modeling dependence of random variables and enable us to separate the marginals from any joint distribution in order to extract the dependence structure. The objective of this paper was to examine the applicability of a copula-CVaR framework in portfolio optimization compared to the traditional MPT. Further, we studied how the presence of memory, when calibrating the copulas, affects portfolio optimization. The marginals for the copula based portfolios were constructed using Extreme Value Theory and the market risk was measured by Conditional Value at Risk. We implemented a dynamic investing strategy where the portfolios were optimized on a monthly basis with two different length of rolling calibration windows. The portfolios were backtested during a sample period from 2000-2016 and compared against two benchmarks: Markowitz portfolio based on normally distributed returns and an equally weighted, non optimized portfolio. The results demonstrated that portfolio optimization is often preferred compared to choosing an equally weighted portfolio. However, the results also indicated that the copula based portfolios do not always beat the traditional Markowitz portfolio. Furthermore, the results indicated that the choice of length of calibration window affects the selected portfolios and consequently also the performance. This result was supported both by the performance metrics and the stability of the estimated copula parameters.

Keywords: Copula, Portfolio Optimization, Extreme Value Theory, CVaR Optimization.
Sammanfattning


Acknowledgement

We would like to express our deepest appreciation to our supervisor Lisa Hed for the useful comments and remarks that you have contributed to this paper. For the theoretical parts of this paper, we would like to tribute five authors in the area of copulas and financial mathematics. Attilio Meucci, thank you for your exceptional papers about copulas and their application in finance. Your ability to explain complex statistical methods in a simple manner has helped us during our learning process. Paul Embrechts, Alexander J. McNeil and Rüdiger Frey, your excellent writings about copulas have given us deep knowledge about the mathematical behind copulas. Last but not least, Carol Alexander, your comprehensive books about financial mathematics and especially the part about copulas has made us understand how several theoretical parts in financial risk modelling are interconnected.

Stockholm, August 31, 2016

Joakim Blom

Joakim Wargclou
“And the wonderful thing about copulas is that the distributions of the marginals can all be different, and different from the copula.”

— Carol Alexander
## Contents

1 Introduction ........................................ 1  
   1.1 Background ...................................... 1  
   1.2 Problem statement .............................. 3  
   1.3 Approach and outline ......................... 4  

2 Theory ................................................. 5  
   2.1 Probability theory ............................ 5  
      2.1.1 Density and distribution functions ....... 5  
      2.1.2 Quantile transformation of a random variable .... 10  
   2.2 Behaviour of financial returns ............... 12  
   2.3 Dependence ..................................... 15  
      2.3.1 Covariance ............................... 15  
      2.3.2 Correlation ............................... 16  
      2.3.3 Copulas .................................. 21  
   2.4 Risk metric .................................... 32  
   2.5 Estimation of the marginal distributions .... 34  
   2.6 Portfolio optimization ....................... 38  

3 Method ................................................. 42  
   3.1 Implementation and software usage .......... 42  
   3.2 Data .......................................... 44  

4 Results ............................................... 46  
   4.1 Summary statistics ............................ 46  
   4.2 Backtesting results ........................... 48  

5 Discussion .......................................... 59  

References .............................................. 62  

Appendix A: Illustration of data .................. 66  

Appendix B: Summary statistics of the data .... 69  

Appendix C: Illustration of estimated copula parameters .... 72  

Appendix D: Portfolio weights ..................... 75
List of Figures

1  The number of published articles regarding copulas from 1990 to 2015. . . . . 3
2  Comparison of elliptical distributions in different dimensions. . . . . . . . 8
3  The cdf and pdf of bivariate normally distributed variables. . . . . . . . . 9
4  The 0.1-quantile of a continuous random variable. . . . . . . . . . . . . 10
5  Quantile transformation of random variables. . . . . . . . . . . . . . . . . 11
6  Fit comparison test between the normal distribution and student-t distribution of the S&P500 returns from 1998 to 2016. . . . . . . . . . . . . . . . 13
7  QQ-plot for daily S&P500 returns from 1998 to 2016. . . . . . . . . . . . 14
8  Observations of two random variables with an obvious dependence $x_2 = x_1^2$
even though the correlation is zero, $\rho_{X_1,X_2} = 0$ . . . . . . . . . . . . . . . 18
9  Simulations from two random variables $X_1, X_2 \sim N(0,1)$ with $\rho = 0.8$
but with different dependence (see more in the section about copulas). a)
Gaussian copula, b) Clayton copula. . . . . . . . . . . . . . . . . . . . . . . 19
10 The pdf of three different copulas: Gaussian, Student-t and Clayton. . . . . 23
11 Simulations from three different copulas: Gaussian, Student-t and Clayton. 30
12 Empirical observations and the fitted cdf created by Pareto tail distribution
and Kernel smoothing. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
13 Efficient frontier together with two portfolios: min CVaR and required
return $\geq 0.05\%$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 40
14 Normalized index returns for the backtesting period 2000-2016. . . . . . 46
15 Backtest of the min CVaR portfolios for a rolling window of 500 days. . . . 49
16 Backtest of the min CVaR portfolios for a rolling window of 1250 days. . . 49
17 Backtest of portfolios with $r \geq 0.01$ for a rolling window of 500 days. . . 50
18 Backtest of portfolios with $r \geq 0.01$ for a rolling window of 1250 days. . . 51
19 Backtest of portfolios with $r \geq 0.02$ for a rolling window of 500 days. . . 52
20 Backtest of portfolios with $r \geq 0.02$ for a rolling window of 1250 days. . . 52
21 Backtest of portfolios with $r \geq 0.03$ for a rolling window of 500 days. . . 53
22 Backtest of portfolios with $r \geq 0.03$ for a rolling window of 1250 days. . . 54
List of Tables

1  Statistical characteristics of the sample data from 1995 - 2016. . . . . . . . 47
2  Mean, volatility and covariance for the yearly assets returns between 1995
   and 2016. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
3  Total return for each portfolio, rolling calibration window and required
   expected of return. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
4  Average daily return for each portfolio, rolling calibration window and re-
   quired expected of return. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55
5  The volatility defined by the standard deviation for each portfolio, rolling
   calibration window and minimum required level of return. . . . . . . . . . 56
6  Maximum draw-down, the worst return between two consecutive days for
   each portfolio, rolling calibration window and a required expected of return. 57
7  Number of month with highest return for each portfolio, rolling calibration
   window and a required expected of return. . . . . . . . . . . . . . . . . . . . 57
8  Number of month with lowest return for each portfolio, rolling calibration
   window and a required expected of return. . . . . . . . . . . . . . . . . . . . 58
1 Introduction

1.1 Background

Through all times, the main purpose for investing has been to construct portfolios where the objective is to find the optimal risk-return trade-off. This objective leads us to the portfolio allocation problem: How should a rational investor allocate the wealth between a set of risky assets? The allocation problem can also be stated in a mathematical way by the portfolio optimization problem where the solution is given by portfolio weights. There are different definitions of portfolio optimization and here is one example from Campbell (2011): “Determination of weights of securities in a portfolio such that it best suits a given objective, eg. maximize return for a given risk”.

Harry Markowitz became a pioneer within this area with his popular and famous paper ”Portfolio-Selection” from 1952. With his mean-variance theory, often called modern portfolio theory (MPT), he formalized the concept of diversification and stated that rational investors should seek the perfect mix of assets rather than the perfect asset. MPT uses future estimates of returns, risk and correlation between assets based on historical data and the optimal portfolio is given by either maximizing the expected return given a level of market risk or minimizing the market risk given an expected return. MPT propose that the optimal portfolios lies on a curve called the efficient frontier and as we shall see later, all portfolios lying below the curve are not efficient. Then one could reallocate the portfolio and lower the risk with the same expected return or increase the expected return for the same level of risk. Hence, MPT assumes that investors are risk averse, which means that investors will choose the portfolio with lowest risk given the same expected return.

However, MPT has since it was introduced been criticized for several shortcomings both empirically and theoretically, see for example Michaud (1989). The criticism is usually directed towards the choice of variance as a risk metric, the use of linear correlation as a measure of dependence and the assumption of normally distributed returns. The normal distribution are easy to work with mathematically but in the real world we have observed that returns are in fact not normally distributed. Instead, distributions of returns have been observed to be both asymmetric and fat tailed (Danielsson, 2011, 9 ff; McNeil et al., 2005, 117 ff). The use of variance has also been criticized, both by Markowitz himself (1999) and by others for being a symmetric risk metric. Besides that the use of variance requires financial time series to be elliptical distributed it also punish returns that are both above and under the mean. Since investors are more concerned with returns that are below mean it is argued that returns above the mean should not be penalized.

According to Danielsson (2011, 76) one of the most common alternative risk metrics is Value at Risk (VaR). This metric is also recommended as a standard by the Basel Committee (Alexander, 2008d, 335). VaR can briefly be described as the predicted maximum loss that will occur over a defined time period with a certain probability (Alexander, 2008d, 13-16). Though, there has been a huge debate about the widely use of VaR as a measure of risk since it in some cases can contradicts the theory of diversification. For
a long time it did not exist a consensus about the properties that a risk metrics should fulfill. To clarify the features of a risk metric, Artzner et al. (1999) developed a framework with four desirable properties. Risk metrics that fulfills these properties are called *coherent risk metrics* and since VaR not obeys all properties, other coherent risk metric were proposed. *Conditional Value at Risk (CVaR)*, also called Expected shortfall, Average Value at Risk or Expected tail loss is one example of a coherent risk metric. CVaR can be interpreted as the expected loss that occurs if the VaR is exceeded (McNeil et al., 2005, 44). In 1999, Rockafeller and Uryasev presented a new approach of how to use CVaR in portfolio optimization instead of using the traditional variance as a risk metric and fortunately, with this method we do not need to assume that returns are normally distributed.

In portfolio optimization we also need to consider the dependence between the assets. MPT uses linear correlation which implies that the dependence is constant for all market conditions. However, empirical results have shown that financial time series exhibit higher dependence in financial crises than in calm periods (Danielsson, 2011, 9ff). Correlation is a widely used term in finance literature when talking about dependence but the terms dependence and correlation are not equivalent and are very often mixed up. Some literature use the word correlation as it covers all types of dependence (Embrechts et al., 2002, 2). In fact, correlation is only one of many measures of stochastic dependence and unfortunately when using it, we have to assume that the returns are elliptical distributed. Embrechts et al. (2002, 6-12) states that when the joint distribution is not elliptical, then there are several shortcomings of using linear correlation. When allowing the marginal distributions of returns to be non-normal and different from each other, correlation loses its meaning as a measure of dependence.

Another way to model dependence is by using *copulas*, which means a "link", "tie" or "bond" according to the Cassell’s latin dictionary. Copulas can be explained as functions that join or link multivariate distribution functions to their one-dimensional marginal distribution function (Nelsen, 2006, 1). According to Durante et al. (2009) and Nelsen (2006), the history of copulas goes back to 1940 and 1951 when Hoeffding and Fréchet introduced the concept of linking multivariate distributions to the marginal distribution. Though it was not until 1959 when Abe Sklar obtained a significant result which proved that copulas can be used to explain the dependence for any multivariate distribution. Sklar’s theorem shows the importance and flexibility of copulas when modelling non-elliptical distributions. Even though the concept of copulas has existed for a long time, it was quite recently that copulas became a very popular and an acclaimed method for financial modelling (Embrechts, 2009). Copulas have by many papers and books been used in portfolio optimization with different approaches and techniques. See for example; Alexander (2008b), Kakouris & Rustem (2014) or Clemente & Romano (2004).

As we can see in Figure 1, the number of academic publications regarding copulas have exploded during the last decade. However, copulas have also been criticized by for example Mikosch (2006), who is sceptical to the lack of critical reviews and states that copulas do not contribute to a better understanding of multivariate extremes. Copulas have also
been pointed out as one factor that affected the sub-mortgage crisis in 2008. In fact, the Gaussian copula has even been called "the formula that killed Wall Street".

![Figure 1: The number of published articles regarding copulas from 1990 to 2015.](image)

When conducting a portfolio optimization we need to estimate the future distributions of returns. For this, there are three main different approaches that can be used; the non-parametric, the parametric and the semi-parametric approach. The non-parametric method is also called historical simulation. When using historical simulation one hopes that the history will repeat itself, i.e. the future behaviour of returns have been captured in the past (Alexander, 2008d, 42-44). Thus, with historical simulation we do not make any assumption of the distribution of the returns. In the opposite to historical simulation, the parametric approach assumes that returns follows some kind of distribution. In finance, one common assumption is to use the normal distribution as in MPT. The semi-parametric model, is as its name suggests, a combination of the parametric and nonparametric approach. The estimation of the future distributions is based on historical data but it is also modelled with some statistical method. Extreme Value Theory (EVT) and Kernel smoothing methods are two examples of methods that can be used under this approach. Kernels smoothing method is a technique for smoothing out the empirical distribution. EVT on the other hand, is a method for handling extreme events by modelling the tail distribution. This method extrapolates the tail distribution in order to predict future extreme events that have not yet occurred. These two methods can also be combined where Kernel smooths the center of the distribution and EVT extrapolates the tails.

### 1.2 Problem statement

As stated above, modern portfolio theory is based on several assumptions that have been theoretically criticized and also showed not to be consistent with the real world stock markets behaviour. For example, MPT measures the risk by variance, assumes that returns are normally distributed and uses linear correlation as a measure of dependence. In this paper we aim to implement a portfolio optimization framework that better capture the non-normality of market returns.
The focus of this paper is to study the applicability of a copula-CVaR framework in portfolio optimization and compare it to the traditional MPT. The comparison will be based on different performance metrics by backtesting the optimal portfolios under a dynamic approach where we reoptimize the portfolio weights on a monthly basis during 2000-2016. Moreover, we will also study how the use of different length of calibration windows, when calibrating the copulas, will affect the optimized portfolios.

1.3 Approach and outline

The outline of the paper is as follows. In chapter two, we describe the theory behind the portfolio optimization used in this paper. We begin with probability theory and financial time series behaviour and continue with dependence structure, risk metrics and portfolio optimization theory. The main focus lays on the sections about modelling dependence with copulas, the semi-parametric approach and the portfolio optimization theory. These parts are of great importance for the reader to fully understand the theories behind our results. In chapter three, we explain the method and data used in this paper. Chapter four contains our results and finally in chapter five, we summarize and discuss our results as well as propose possible extensions for further research in this area.
2 Theory

2.1 Probability theory

To conduct a portfolio optimization we need to model the future distributions of returns. This requires some basic knowledge about probability and statistics and consequently, we start to define the properties of univariate and multivariate distributions which also prepares basis for the understanding of copulas.

2.1.1 Density and distribution functions

We start to define a probability space, which is used to describe an experiment in a mathematical way. A probability space, which is constructed with respect to a specific event consists of three parts: $(\Omega, \mathcal{A}, \mathbb{P})$, where $\Omega$ is the set of all possible outcomes, $\mathcal{A}$ (also called sigma-algebra) is the collection of events where the events have zero or more outcome and $\mathbb{P}$ is the probability measure (Kloeden and Platen, 1999, 3).

A random variable (r.v.) is a variable whose values are stochastic which means that there is uncertainty about the values that the variable can take. In financial modelling, returns on financial assets are considered to be r.v.s. Every outcome of the r.v. is determined by a chance event but we can measure the probability of which value it will take. The set of all outcomes and their related probabilities is called a probability measure (Alexander, 2008a, 75). In a mathematical way, given a probability space, a r.v. $X$ is defined by: $X : \Omega \rightarrow \mathbb{R}$ (real-valued), if it is $\mathcal{A}$-measurable. For the coming section, we assume that the reader is familiar with the basic laws of probability and statistics and content ourselves to remind that the probability of an event $A$ is a number between 0 and 1 given by the notation: $\mathbb{P}(A) \in [0, 1]$. For the understanding of this paper we also state that the a random variate is a particular outcome of a r.v..

Univariate distributions

Two different ways to represent a probability measure of a r.v. are either by the probability density function or by the probability distribution function which is also called the cumulative distribution function. Every distribution is characterized by the four moments: the expected value (also called expectation), variance, skewness and kurtosis. These moments determines the shape and behaviour of the distributions. The first two, expectation and variance describes the location and dispersion of the distribution whereas skewness determines the symmetry and the kurtosis measures whether the data are heavy-tailed. See Alexander (2008a, 78-83) or Tsay (2005, 8-9) for mathematical definitions.

For the following section, we let $X$ be a continuous r.v. with a given domain $D$ in $\mathbb{R}$. For example, $D$ can be $(-\infty, \infty)$ or $[0, \infty)$. We also assume the distributions are continuous because when we introduce copula distributions this assumption will become necessary. Whenever details are missing we refer the reader to Alexander (2008a, 85 ff) or Tsay (2005, 7 ff)
The probability density function (pdf), $f_X$, of a r.v. is an integrable, non negative function such that:

$$
\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx.
$$

Further, the total area under the curve must be 1 since $\mathbb{P}(-\infty \leq X \leq \infty) = 1$, and for continuous r.v.s. we also have that $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b)$. Given the pdf $f_X$, the corresponding cumulative distribution function (cdf) $F_X$ is a continuous increasing function that maps $F_X : D \rightarrow [0, 1]$ such that:

$$
F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_X(u) du.
$$

Then for any $x \in D$, the value of the distribution function $F_X$ at $x$ is the probability that $X$ is less or equal than $x$. The cdf also fulfills $F_X(-\infty) = 0$ and $F_X(\infty) = 1$. The relationship between the pdf and cdf is given by $F_X = \frac{d}{dx} F_X$. If the cdf $F_X$ is strictly increasing and continuous, then the inverse distribution function (also called the quantile function) is a function that maps $F_X^{-1} : [0, 1] \rightarrow D$ and is given by:

$$
F_X^{-1}(u) = F_X^{-1}(F_X(x)) = x, \quad \forall x \in D, \ u \in [0, 1].
$$

Another important distribution function for the understanding of copulas is the standard uniform distribution. This distribution is given by the notation $X \sim [0, 1]$ and means that a r.v. $X$ can take any value between 0 and 1 with the same probability.

**Multivariate distributions**

A multivariate distribution contains more than one r.v. that might be dependent of each other. Often in this paper we consider a bivariate distribution, i.e. a distribution function given in two dimensions, since it is easier to illustrate. Let $X$ and $Y$ be two continuous r.v.s. defined on a domain $D \subset \mathbb{R}$. Then the bivariate density function $f_{X,Y}$ in $\mathbb{R}$, is an integrable, non negative function with a total area under the curve that equals 1, such that:

$$
\mathbb{P}(x_a < X < x_b, y_a < X < y_b) = \int_{y_a}^{y_b} \int_{x_a}^{x_b} f_{X,Y}(x,y) \, dx \, dy.
$$

The bivariate density function returns the joint probability that $X$ and $Y$ takes certain values $x, y \in D$ simultaneously and the shape of the density function depends on the marginal densities for $X$ and $Y$ and the dependence between $X$ and $Y$.

If $f_{X,Y}$ exist, the corresponding bivariate joint distribution is a continuous non decreasing function denoted by $F_{X,Y}$, that maps $F : D^2 \rightarrow [0,1]$. If $F_{X,Y}$ is partially differentiable this function is given by:

$$
F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y) \, dy \, dx.
$$
This can easily be extended to $d$ dimensions. Let $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ be $d$ r.vs. with respectively marginal distributions $F_{X_1}, F_{X_2}, \ldots, F_{X_d}$. Then the joint distribution is given by:

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = \mathbb{P}(X_1 < x_1, \ldots, X_d < x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{\mathbf{X}}(x_1, \ldots, x_d) \, dx_1, \ldots, dx_d.$$ 

Given that the marginal distributions are continuous, the multivariate pdf is given by differentiating the cdf with respect to all r.vs.:

$$f_{\mathbf{X}}(x_1, \ldots, x_d) = \frac{\partial^d F_{\mathbf{X}}(x_1, \ldots, x_d)}{\partial x_1, \ldots, \partial x_d}.$$ 

The cdf is also known as a marginal distribution if the function $F_{\mathbf{X}}$ is given in one dimension. Consider once again the vector $\mathbf{X}$ of r.vs. with respectively marginal distribution $F_{X_1}, F_{X_2}, \ldots, F_{X_d}$. Then each marginal distribution $F_{X_i}$ alone ignores any information about the other marginals. One marginal can be obtained from the joint distribution by integrating out $X_1$ from $\mathbf{X}$:

$$F_{X_1}(x) = F_{\mathbf{X}}(x_1, \infty, \ldots, \infty).$$

The marginal densities are in the same way given by:

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, \ldots, x_d) \, dx_2, \ldots, dx_d.$$ 

The r.vs. in $\mathbf{X}$ are said to be independent if the outcome of $X_i$ contains no information about the outcome of the others. In mathematical meaning, we say that the r.vs. are independent if the joint distribution is a product of the marginal distributions:

$$f_{\mathbf{X}}(x_1, \ldots, x_d) = f_{X_1}(x_1) \cdots f_{X_d}(x_d).$$

Observations from a bivariate independent distribution would look like the two dimensional uncorrelated data in Figure 2.
Elliptical distributions

Elliptical distributions are a generalisation of multivariate normal distributions and the easiest way to illustrate an elliptical distribution is by using a bivariate distribution formed as an ellipse. In Figure 2 below we can see elliptical distributions in different dimensions. In 1-dimension, elliptical symmetry means that the marginal distribution is symmetric.

![Figure 2: Comparison of elliptical distributions in different dimensions.](image)

Below we define two of the most common elliptical distributions: the multivariate normal- and Student-t distribution. For an explicit mathematical definition of elliptical distributions and their properties, see McNeil et al. (2005, 89 ff). These two distributions are presented since a common assumption in financial risk modelling is that returns are multivariate normally distributed and later on we will present why linear correlation only is applicable when considering elliptical distributions.

Let $\mathbf{x}$ be a $k$-dimensional vector of returns, $\mathbf{\mu}$ be a vector of expected returns and $\mathbf{\Sigma}$ is a symmetric, positive definite covariance matrix with the determinant $|\mathbf{\Sigma}|$. Then, the
multivariate normal density function is given by:

\[
f(x) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} \exp \left( -\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu) \right).
\]

An important property is that the portfolio return will be normally distributed if the returns for all assets included in the portfolio are normally distributed. This is because a linear transformation of a multivariate normal random vector also has a multivariate normal distribution.

A generally more sophisticated distribution to use is the Student-t distribution. The Student-t distribution is linked to the normal distributions since it is symmetric but differs since it has heavier tails which allows for more extreme values. The Student-t distribution is, except for the correlation matrix, described by the parameter "degrees of freedom" \(\nu\), which determines the length of the tails and peak of the center. A higher number of degrees of freedom increases the peak of the center and lowers the tails. The Student-t becomes closer and closer to the normal distribution as \(\nu\) increases.

Given a \(k\)-dimensional vector of returns \(x\) with corresponding expected returns \(\mu\), a covariance matrix \(\Sigma\) and \(\nu \in \mathbb{R}_+\) the degrees of freedom. Then the probability density function of the multivariate Student-t distribution is given by:

\[
f(x) = \frac{\Gamma[(\nu + k)/2]}{\Gamma(\nu/2)\nu^{k/2} \pi^{k/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(x - \mu)^T \Sigma^{-1} (x - \mu)\right]^{-\nu+k/2},
\]

where \(\Gamma(\cdot)\) is the gamma function, see Alexander (2008b, 118) for expression.

![Figure 3: The cdf and pdf of bivariate normally distributed variables.](image)
2.1.2 Quantile transformation of a random variable

Simulation of a r.v. using the quantile function is a very important concept to understand before reading the section about copulas later on. Let $X$ be a continuous r.v. with a domain $D$ and associated with a probability $\alpha \in [0, 1]$. Then, the $\alpha$ quantile of $X$ is the value $x_\alpha$ of $X$ such that $\Pr(X < x_\alpha) = \alpha$. The 0.1-quantile of a continuous r.v. is illustrated in Figure 4 below.

![Probability density function (pdf) and Cumulative density function (cdf)](image)

*Figure 4: The 0.1-quantile of a continuous random variable.*

Simulation of a r.v. using a quantile is given by the following algorithm:

**Algorithm 1:** Simulations using quantiles

1. Simulate a random number $u$ from the standard uniform distribution $U \sim U[0, 1]$ to represent a probability.

2. Use the inverse distribution function, $F_X^{-1}$, to find the corresponding quantile:

   \[ x = F_X^{-1}(u), \]

   where $x$ is the $u$ quantile of $X$.

We try to clarify the theory behind the simulation above by the following explanation. Consider a standard uniform variable $u$ which has the property $\Pr(U < u) = u$, because uniform variables have linear distribution functions. Hence, we have that for all $u \in [0, 1]$:

\[ \Pr(F_X(X) < u) = \Pr(X < F_X^{-1}(u)) = F_X(F_X^{-1}(u)) = u. \]

This means that we can transform a r.v. $X$ to a standard uniform variable $F_X(X) \sim U[0, 1]$. 
by applying the cdf (Alexander, 2008b, 258-259; Meucci, 2011, 3-4)). In the opposite way, we can also transform a variable from the standard uniform distribution to simulate from the distribution of X. This transformation is conducted by letting \( u = F_X(x) \) such that \( P(U < F_X(x)) = P(F_X^{-1}(U) < x) \). Thus, we can simulate a set of independent r.v.s. from the distribution of X by simulating r.v.s. \( u \) from \( U \sim U[0, 1] \) and apply the inverse distribution function to obtain the corresponding quantile for X. When simulating several r.v.s. we have to take their dependence into account and this will be explained in more detail in section 2.3.3 about copulas. The transformation of r.v.s. using the inverse distribution function is shown in Figure 5 below.

\[ \text{Figure 5: Quantile transformation of random variables.} \]

Stochastic processes
Before we move on to the sections about dependence and risk metrics, we need to put the r.v.s. in terms of a stochastic process. This theory is not used in practice when modelling the dependence with copulas. However, it is needed when we later on consider the simulated returns as independent in time but dependent of each other.

A stochastic process is a sequence of identically distributed r.v.s., \( X_1, X_2, \ldots, X_d \) denoted by: \( \{X_t\}_{t \in T} \), where \( t \) states a point in time. It is common to consider the r.v.s. to be continuous but the sequence may be over both continuous or discrete time. There are different classes of stochastic processes and one important type is the stationary process. According to Alexander (2008a, 134-135) a discrete stochastic process is stationary if:

- The expected value, \( E(X_t) \), is finite and constant.
- The variance, \( V(X_t) \), is finite and constant.
Theory

- Given two points in time $s, t$, the joint distribution for $(X_s, X_t)$ depends only on $|t - s|$.

In other words, the random variables $X_t$ of the stationary process $X$ have the same distribution for all $t \in T$ (Kloeden and Platen, 1999, 28).

One common assumption is to consider the r.vs. of a stationary stochastic process to be independent and identically distributed (i.i.d.). The assumption for independence implies that there is no autocorrelation in the process, i.e. there is no dependence in time. The assumption of identical distributions means that the r.vs. all have the same distribution parameters for all time steps $t \in T$. In practise it is often hard to prove that time series are stationary and as a result, the third condition above is often weakened so that only the covariance is independent of time:

$$\text{Cov}(X_s, X_t) \text{ dependence only on } |t - s|.$$ 

This process is then called a weakly stationary or covariance stationary process (Danielsson, 2011, 209). In this paper we will assume that the returns follows a weakly stationary process with i.i.d. r.vs.

2.2 Behaviour of financial returns

Statistical studies of financial time series have showed that returns tend to follow some statistical properties. According to Danielsson (2011, 9 ff), these three properties are called the stylized facts of financial returns:

- Volatility clusters
- Fat tails
- Nonlinear dependence

Volatility clustering of certain speculative prices was first observed by Mandelbrot (1963) who noted that: “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.” Mandelbrot’s observation of volatility clustering has inspired many financial researchers in the development of stochastic models in order to capture this phenomenon more accurately. Examples of some of the common models in this area are ARCH (Engle, 1982) and GARCH (Bollerslev, 1986). Nearly all financial time series exhibits volatility clustering and in a mathematical sense this means that the standard errors do not follow an i.i.d stochastic process. This can also imply that the standard errors have autocorrelation and/or heteroscedasticity (Alexander, 2008a, 184). Autocorrelation means that the return on one day is correlated with the return on a previous day. Heteroscedasticity means that the variance is not constant but do not automatically imply autocorrelation. However, in this paper we will not consider volatility clusters in our risk model since the major aim is to study dependence between assets rather than dependence in time. Though, models that handles both autocorrelation and heteroscedasticity as well as dependence between assets is one obvious improvement.
for further research within portfolio optimization.

The second stylized fact, fat tails, points out the fact that the distribution of returns do not follow a normal distribution. The normal distribution is determined by the first two moments of a r.v., mean and variance. This implies that the third and fourth moments, skewness and kurtosis are the same for all normally distributed variables. As mentioned in a previous section, skewness is a measure of the asymmetry of a distribution while kurtosis measures the degree of fat tails relative the peak of a distribution. According to empirical observations, extreme fluctuation in financial returns occurs more likely compared to a normal distribution.

If we want to test our data for fat tails, we can either do this by graphical methods or with statistical tests. Two different graphical methods are QQ-plot and histogram. By using a histogram we can illustrate the sample density distribution compared to the normal distribution. In Figure 6 we see a histogram of returns together with a normal fit and a Student-t fit. Since the Student-t distribution is a fat-tailed distribution, it will often approximate the distribution of returns more accurate compared to the normal distribution. The QQ-plot (quantile-quantile plot), is another graphical tool to help us assess if a set of data plausibly comes from some theoretical distribution such as a normal distribution. By using a QQ-plot with different degrees of freedom, $\nu$, we obtain a graphical illustration of how fat the tails of our distribution actually are. As we see from Figure 7, in this case the Student-t distribution with 3 degrees of freedom approximates the returns more accurate compared to the normal distribution.

![Figure 6: Fit comparison test between the normal distribution and student-t distribution of the S&P500 returns from 1998 to 2016.](image)
The two most common statistical methods to test if the data follows a normal distribution are the Jarque–Bera test (JB-test) or the Kolmogorov–Smirnov test (KS-test). In this paper we use the JB-test as a complement to the QQ-plot but we do not go into any further details about the test. Instead see Alexander (2008a, 158) or Danielsson (2011, 16-17) for more details.

The final stylized fact is nonlinear dependence. In financial risk modelling it is common to assume a linear dependence structure between assets which implies that the dependence is the same in all market conditions. Though, research has shown that dependence between financial assets tend to be higher in market turbulence rather then in calm periods. In fact, in a financial crisis assets tend to be fully dependent (Danielsson, 2011, 21-24). Nonlinearity for a sample of returns can be illustrated by calculating and analyzing the consistent of correlation over time.

These properties above are the main reason why we turn to model the dependence with copulas instead of traditional linear correlation. As we shall see in a later section, there are several shortcomings when using linear correlation and copulas enable us to consider the first two stylized facts and model dependence in a more realistic way.

Figure 7: QQ-plot for daily S&P500 returns from 1998 to 2016.
2.3 Dependence

In the previous section we have presented the basics about probability theory and behaviour of financial returns. As we have seen, researchers have presented empirical evidence that there exist fat tails and nonlinear dependence between financial returns. In this section, we will go further into how to model dependence between financial time series. But before we explain copulas, we start with presenting and explaining covariance and correlation to see why this type of dependence is not applicable when considering non normal distributions.

2.3.1 Covariance

To define correlation we have to start with covariance, which is a scalar measure of linear dependence. Let $X$ and $Y$ be two r.vs., then the covariance is given by:

$$\sigma_{X,Y} = Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)],$$

where $\mu_x = E(X)$ and $\mu_y = E(Y)$ are the expected values of $X$ and $Y$ (Embrechts, 2002, 6). The variance of one r.v., for example $X$, is given by the notation $V(X) = Cov(X, X)$. In the multivariate case, consider vectors of r.vs. $X = (X_1, \ldots, X_d)$ in $\mathbb{R}^d$. All pairwise covariances can then be summarized in a $d \times d$ matrix, $Cov(X)$. This matrix will be symmetric, squared and at least positive semi-definite with variances along the diagonal and covariances as the non diagonal elements. Given a symmetric matrix $A$ of size $d \times d$, then $A$ is a positive semi-definite matrix if $x^T A x \geq 0$, for every vector $x$. In the opposite way, $A$ is negative semi-definite if $x^T A x \leq 0$ for every vector $x$.

In practice we do not often know the covariance matrix since we work with samples of observations and therefore we need to estimate it. Consider again the two r.vs. $X$ and $Y$ with $t$ observations $x_1, \ldots, x_t$ and $y_1, \ldots, y_t$. In most cases when we have to estimate the means by $\bar{x}$ and $\bar{y}$, the sample covariance, $\hat{\sigma}_{X,Y}$, is according to Danielsson (2011, 203) given by:

$$\hat{\sigma}_{X,Y} = \frac{1}{t-1} \sum_{i} (x_i - \bar{x})(y_i - \bar{y}).$$

Unfortunately is the covariance hard to interpret since the magnitude of covariance depends upon the units by which the r.vs. are measured (Alexander, 2008a, 112). Another shortcoming is that covariance only measures linear dependence and works for elliptical distributions. This will be explained in more detail in the coming section about Pearson’s correlation.
2.3.2 Correlation

Correlation, which is a standardized measure of strength of dependence between r.vs., is closely linked to covariance. There exist different types of correlation measures but the most commonly used is Pearson’s correlation.

Pearson’s correlation

Pearson’s correlation is the most widely used among the measures of correlation. But according to Embrechts et al. (2002, 6), the term is often presented as a measure that would cover all types of dependence when in fact, Pearson’s correlation is just one of many measures of dependence between r.vs. If we consider two real-valued r.vs. $X$ and $Y$ with finite variances, then Pearson’s correlation is given by:

$$ \rho_{X,Y} = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}, $$

where $\text{Cov}(X, Y)$ is the covariance between $X$ and $Y$ and $V(X), V(Y)$ are the variance for $X$ and $Y$ respectively. Like covariance, Pearson’s correlation is a scalar measure of linear dependence which means that it takes a single value, but since it is normalized the range are given by $[-1, 1]$. If the correlation is 1 it implies a perfect positive correlation and in the opposite, -1 implies a perfect negative correlation. When the r.vs. are independent the correlation is zero, but as we explain later, the opposite does not always hold.

There are several reasons why Pearson’s correlation is so popular to use. One reason is that we only have to calculate the first two moments, mean and variance, of the distribution before calculating the correlation coefficient. Consider again the two r.vs. $X$ and $Y$ with $t$ observations $x_1, \ldots, x_t$ and $y_1, \ldots, y_t$. Then Pearson’s correlation can be estimated by the sample correlation coefficient $\hat{\rho}_{X,Y}$:

$$ \hat{\rho}_{X,Y} = \frac{\sum_{i=1}^{t}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{t}(x_i - \bar{x})^2 \sum_{i=1}^{t}(y_i - \bar{y})^2}}, $$

where $\bar{x}$ and $\bar{y}$ are the sample means of $X$ and $Y$, given by: $\bar{x} = \frac{1}{t} \sum_{i=1}^{t} x_i$ and in the same way for $\bar{y}$ (Danielsson, 2011, 203). In the multivariate case with $d$ assets we can, in the same way as with covariance, summaries all pairwise correlation coefficients in a $d \times d$ matrix. This makes it easy to illustrate the linear dependence among several assets.

According to Tierens and Anadu (2004) we can also estimate the average stock correlation for a portfolio $P$ consisting of $d$ assets with portfolio weights $w_i$, $i = 1, \ldots, d$, by the following equation:

$$ \hat{\rho}_P = \frac{2 \sum_{i=1}^{d} \sum_{j>i}^{d} w_i w_j r_{i,j}}{1 - \sum_{i=1}^{d} w_i^2}. \tag{1} $$

We will use this formula to evaluate the calibrated correlation matrix for the copula functions, which will be described in more detail later on. It is important to understand that Pearson’s correlation only works well when dealing with elliptical distributions such as
the multivariate normal distribution (McNeil et al., 2005, 201 ff; Alexander, 2008b, 95-96). We also need to assume that the underlying data comes from a stationary stochastic processes, such as the i.i.d. process when we use Pearson’s correlation (Alexander, 2008a, 111). As stated above, it is not often real world financial data satisfies these assumptions and the fact that Pearson’s correlation suffers from being a linear measure of dependence has been illustrated in the excellent papers by Embrechts et al. (2002) and by McNeil et al. (2005) both in terms of shortcomings and fallacies. The next section describes some of these shortcomings when using Pearson’s correlation in financial risk modelling. These are presented as we believe it creates a better understanding of why we choose to model the dependence with copulas. For a more detailed description of the shortcomings we refer the reader to the mentioned paper.

Shortcomings and fallacies with Pearson’s correlation
Consider again the case of two real-valued r.vs. $X$ and $Y$. Then the following weaknesses with Pearson’s correlation have been identified:

- Zero correlation between the r.vs. $X, Y$ does not imply that the variables are independent. Though, on the other hand if the r.vs are independent it implies that they have zero correlation, $\rho_{X,Y} = 0$. Embrechts et al. (2002, 7-8) explains that only when dealing with multivariate normal distributions it is acceptable to interpret zero correlation as implying independence.

  This shortcoming can be considered by the following example where correlation does not exist although a strong dependency appears. Let $X \sim N(0, 1)$ be a r.v. from the standard normal distribution. Further if we let $Y = X^2$ we have that $\rho_{X,Y} = 0$, even though a clear dependence appears. This is illustrated in Figure 8 below.

- The linear correlation will only be defined if the variances of $X$ and $Y$ are finite. According to Embrechts et al. (2002, 7) and McNeil et al. (2005, 202) this restriction can become troublesome when working with distributions having fat tails.

- Under nonlinear strictly increasing transformations, linear correlation is not invariant: $T: \mathbb{R} \rightarrow \mathbb{R}$. This means in general, that for two real-valued r.vs. we have:

  $$\rho(T(X), T(Y)) \neq \rho(X, Y).$$

  A nonlinear transformation can be explained as a transformation that changes the relation between the r.vs and this means that the correlation will change.
We now continue to present further pitfalls, presented in Embrechts et al. (2002, 21ff) and McNeil et al. (2005, 202-205), in terms of fallacies. These papers state that the fallacies are of great importance to know since they bring attention to danger when constructing multivariate distributions based on correlation and marginal distributions.

- **Fallacy 1:** "The marginal distributions and pairwise correlations of a random vector determine its joint distribution."

  This statement will be true if we restrict ourselves to elliptical distributions. Consider two r.vs. $X$ and $Y$ with correlation $\rho_{X,Y}$. If the marginal distributions and the correlation are unknown there are many possible bivariate distributions for $(X,Y)$. But if we instead know that the joint distribution is a bivariate normal distribution, then the correlation and marginals uniquely determines the joint distribution. This fallacy can be demonstrated by Figure 9 where we illustrate two different bivariate distributions, both with correlation $\rho_{X,Y} = 0.8$. It is clear that there is a more risky dependence in the lower tail in b) than in a).

- **Fallacy 2:** "Given marginal distributions $F_X$ and $F_Y$ for the r.vs. $X$ and $Y$, all linear correlation, $\rho_{X,Y}$, in [-1 1] can be attained by suitable specifications of the joint distribution."

  As with fallacy 1, this is true if the marginals, $F_X$ and $F_Y$ are elliptical distributions, but in general it is false. The possible values of correlation can be given in a subset to $(-1,1)$, see McNeil et al. (2005, 203-205) for a proof and example.

- **Fallacy 3:** "Let two risky assets be given by the r.vs. $X$ and $Y$. Then Value-at-Risk (VaR) for the sum of the two assets $X + Y$ is at worst when the correlation, $\rho_{X,Y}$ is maximal, or in other words, when $X$ and $Y$ are comonotonic."

  Again, this statement is true if we are dealing with elliptical distributions but in general it is false. We will discuss this problem in a later section but for a numerical
proof, see McNeil et al. (2005, 251).

For the understanding of fallacy 3, we shortly explain the terms *monotonic* and *comonotonic*. A monotonic function can be explained as a function which is either entirely nondecreasing or nonincreasing. Or in a mathematical sense we can say that if the first derivative of a function does not change sign, then the function is monotonic. Variables with perfect positive dependence, i.e. strictly increasing are called *comonotonic* and variables with perfect negative dependence, i.e. strictly decreasing are called *countermonotonic*. For a more formal definition of comonotonicity and countermonotonicity, see McNeil et al. (2005).

An important meaning from this section is that interpretations from Pearson’s correlation should be made with caution and can almost be seen as meaningless unless it is applied to a suitable distribution.

![Figure 9: Simulations from two random variables $X_1, X_2 \sim N(0,1)$ with $\rho = 0.8$ but with different dependence (see more in the section about copulas). a) Gaussian copula, b) Clayton copula.](image)

**Concordance metrics**

Before moving on to other measures of dependence, we need to specify the basic properties that a good measure of association between r.vs. should have. In the coming section we introduce the concept of concordance and two fundamental concordance metrics, Spearman’s rho and Kendall’s tau, that later on will be shown to have an important link with copulas. Consider two points $(x_1, y_1)$ and $(x_2, y_2)$ in $\mathbb{R}^2$. The points are said to be:

- **concordant**, if $(x_1 - x_2)(y_1 - y_2) > 0$.

- **discordant**, if $(x_1 - x_2)(y_1 - y_2) < 0$.

Two examples of desired properties that a concordance metric should fulfill are given in Alexander (2008b, 255-256) and in Embrechts et al. (2002, 15-16) respectively. The properties we choose to present below are taken from the former. Let $\delta$ be a measure of dependence which assigns a real number to any pair of real-valued r.vs. $X$ and $Y$. Then,
a measure of concordance, $\delta(X,Y)$, is a numerical measure of association between two continuous r.vs. $X$ and $Y$ that satisfies the following properties:

1. Symmetry: $\delta(X,Y) = \delta(Y,X)$ and $\delta(X, -Y) = -\delta(X,Y)$.
2. Normalisation: $-1 \leq \delta(X,Y) \leq 1$.
3. Comonotonicity and countermonotonicity: $\delta(X,X) = 1$, $\delta(X,-X) = -1$.
4. Independence: $X$ and $Y$ are independent $\Rightarrow \delta(X,Y) = 0$.
5. Consider two different joint distributions given by the notation $F_{X,Y}$ and $G_{X,Y}$. Further let $\delta_F(X,Y)$ and $\delta_G(X,Y)$ denote concordance measures under the two distributions. Then, given that $F_{X,Y}(x,y) \geq G_{X,Y}(x,y)$ we must have $\delta_F(X,Y) \geq \delta_G(X,Y)$.

Noteworthy is that Pearson’s correlation only fulfills the first and second property above. Rank correlation, which will be presented in a coming section, fulfills all properties above if $X$ and $Y$ are continuous. It is important to note that this is only one selection of properties and they can be extended or revised if desired. We will now move on and provide two examples of concordance metrics that have an important role when calibrating copulas.

**Rank correlation**

Rank correlation is another scalar measure of dependence between r.vs.. Unlike Pearson’s correlation, rank correlation does not make any assumption of the distribution other than that the r.vs. should be monotonic. There are two main types of rank correlation: Kendall’s tau, $\tau$, and Spearman’s rank correlation, $\rho_s$. Roughly explained, Pearson’s correlation measures linear dependence while the rank correlations measures strictly monotonic dependence (Nelsen, 2006, 157). The standard estimators of rank correlation are calculated based on the ranked data which means that we do not need to know the actual values. The definitions of the rank correlations below can be found in Embrechts et al. (2002, 16).

**Spearman’s rank correlation**

Let $X$ and $Y$ be two r.vs. with distribution functions $F_X$ respectively $F_Y$ and let the joint distribution be $F_{X,Y}$. Then, the Spearman’s rank correlation, $\rho_s$ is given by:

$$\rho_s(X,Y) = \rho(F_X(X), F_Y(Y)),$$

where $\rho$ often is Pearson’s correlation. To estimate Spearman’s rank correlation, we rank the data for each of the two vectors individually and simply calculate Pearson’s correlation on the ranked data. In mathematical notations, consider a sample of $d$ paired observations $(x_1, y_1), \ldots, (x_d, y_d)$, where $\text{Rank}(x_i)$ and $\text{Rank}(y_i)$ denotes the rank of each variable. Spearman’s rank correlation coefficient is then given by:

$$\rho_s = 1 - \frac{6K}{d(d^2 - 1)},$$
where $K = \sum_{i=1}^{d} (\text{Rank}(x_i) - \text{Rank}(y_i))^2$.

**Kendalls Tau**

Another rank correlation is Kendall’s tau. Let $(X_1, Y_1)$ and $(X_2, Y_2)$ be two independent pairs of r.vs. from the joint distribution $F_{X,Y}$. Kendall’s rank correlation is then given by:

$$\rho_\tau(X,Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

By the definition above, Kendall’s rank correlation is the probability of concordance minus the probability of discordance for all pairs.

Let $(x_i, y_i), i = 1, \ldots, d$, be a sample of $d$ observations with all possible pairs given by $\binom{d}{2} = \frac{d(d-1)}{2}$. Further we let $N_C$ be the number of concordant pairs and $N_D$ be the number of discordant pairs. Kendall’s sample tau, $\tau$ is then estimated by:

$$\tau = \frac{N_C - N_D}{\frac{d(d-1)}{2}}. \quad (2)$$

Spearman’s rank correlation and Kendall’s tau do not give the same values except for some special joint distributions, but the two can be related to each other, see Nelsen (2006) for the relationship. For the definitions of the rank correlation estimators we refer the reader to Alexander (2008b, 256). As we shall see later, we use rank correlation to calibrate copulas to fit the empirical data.

**2.3.3 Copulas**

In this section we will present the theory needed to understand how we model the dependence between returns with copulas. Before moving on to definitions and properties of copulas, we recall the important concept of quantile transformations of r.vs. presented in section 2.1.2. We recommend the reader to fully understand that concept before reading about copulas.

By using copulas we can construct a function that only describes the dependence structure. In detail, given any joint distribution, copulas enable us to separate the dependence from the marginals. If we consider a random vector of risk factors, for example historical returns from a portfolio of stocks. Then this vector contains both information about the behaviour of the individual stocks, given by the marginal distributions, but also about the dependence between the stocks. When using copulas we can extract this dependence in form of a distribution function to analyze and construct new joint distributions. Copulas are a very useful tool for modelling complex dependence structures such as when the marginals are not assumed to be normally distributed. For instance, we could let the returns on one stock be given by a gamma distribution whilst the returns on the other stock are given by a Student-t distribution. Such a case requires the use of copulas since linear correlation will be meaningless.
Basic properties of copulas.
Consider $d$ r.v.s. $X_1, \ldots, X_d$ with continuous cdfs $F_{X_1}, \ldots, F_{X_d}$ and let $u_i = F_{X_i}(x_i), i = 1, \ldots, d$. A $d$-dimensional copula $C(u) = C(u_1, \ldots, u_d)$ is a distribution function on $[0, 1]^d$ with standard uniform marginal distributions that fulfills the following properties:

1. $C : [0, 1]^d \to [0, 1]$.
2. $C(u_1, \ldots, u_d)$ is increasing in each component $u_i$.
3. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i, \forall i \in \{1, \ldots, d\}$, $u_i \in [0, 1]$.
4. $\forall (a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [0, 1]^d$ with $a_i \leq b_i$ we have:
   $$\sum_{i_1=1}^{2} \ldots \sum_{i_d=1}^{2} (-1)^{i_1+\ldots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,$$
   where, $u_{j1} = a_j$ and $u_{j2} = b_j \forall j \in \{1, \ldots, d\}$.

These properties can be seen in Alexander (2008b, 259-260) and McNeil et al. (2005, 185). The first property states that a copula function $C$ is a joint cumulative distribution function of a $d$-dimensional random vector and the second property is required for any multivariate distribution function. The third property is the requirement of uniform marginal distributions and the final fourth property ensures that if the vector $(U_1, \ldots, U_d)^T$ has the distribution function $C$, then $P(a_1 \leq U_1 \leq b_1, \ldots, a_d \leq U_d \leq b_d)$ is non negative (McNeil et al., 2005, 185). We need at least two dimensions to construct a copula since a copula determines the dependence structure. Hence, for $2 \leq k < d$, the $k$-dimensional distributions of the $d$ dimensions are also copulas.

Sklar’s Theorem
By the definition of copulas, many possible joint distributions can be constructed. But to use copulas we need to know how the copula uniquely links the marginal distributions to the joint distribution. Sklar (1959) formulated a relationship called ”Sklar’s theorem” which shows that all multivariate distributions contains a copula and that joint distributions can be constructed by using a copula and marginal distributions (Nelsen, 2006, 260-261)

**Theorem 2.1. Sklar’s Theorem.**
Let $F$ be a joint distribution function with marginals $F_{X_1}, \ldots, F_{X_d}$. Then, for all $x_1, \ldots, x_d \in \mathbb{R}$, there exists a copula $C : [0, 1]^d \to [0, 1]$ such that:

$$F_X(x_1, \ldots, x_d) = C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)). \quad (3)$$

If the marginal distributions are continuous, then $C$ is unique. Otherwise $C$ is uniquely determined by the product of the range of $F_X$.

On the other hand, if $C$ is a copula and $F_{X_1}, \ldots, F_{X_d}$ are univariate distribution functions, then the function $F_X$ defined in (3) is a joint distribution function with marginals $F_{X_1}, \ldots, F_{X_d}$.
For the proof of existence and uniqueness of a copula when the marginals are continuous, see McNeil et al. (2005, 186-187). The proof shows how joint distributions can be formed by coupling together marginal distributions with a copula \( C \), and how copulas can be extracted from any multivariate distribution with continuous marginal distributions.

One important type of copulas are *implicit copulas* which are extracted from well known multivariate distributions using Sklar’s Theorem. By setting \( F_{X_i} = u_i \) we can construct an implicit copula of any joint distribution \( F_X \):

\[
C(u_1, \ldots, u_d) = F_X(F_{X_1}^{-1}(u_1), \ldots, F_{X_d}^{-1}(u_d)).
\]

(Copulas density function)

Another important concept is called the *copula density function*. According to Alexander (2008b, 260-261) we get the following expression for the copula density function by differentiating (3):

\[
c(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)) = \frac{\partial^d C(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))}{\partial F_{X_1}(x_1) \ldots \partial F_{X_d}(x_d)},
\]

where \( u_i = F_{X_i} \). If the copula density function exist, the joint density can be expressed as:

\[
f_X(x_1, \ldots, x_d) = f_{X_1}(x_1) \ldots f_{X_d}(x_d) \cdot c(F_{X_1}(x_1), \ldots, F_{X_d}(x_d)).
\]

(Copulas density function)

Compared to ordinary density functions which often have the biggest values in the center, copula densities often looks differently. Three common bivariate copula density functions; Gaussian, Student-t and Clayton, are presented in Figure 10. If the variables are independent then the copula density is one everywhere but the most common copulas in finance have higher values in the tails than in the center. This indicates that the dependence are higher when returns are observed in the tails, for example when returns are observed to be high or low.

![Figure 10: The pdf of three different copulas: Gaussian, Student-t and Clayton.](image)

Another important feature with copulas is that they are invariant under strictly increasing transformations of the marginals. This is a desirable property and makes the copulas a good representation of dependence, see Meucci (2011, 7-8) or Nelsen (2006, 4) for examples of this property.
Families of copulas.
In this section we introduce the copulas used in this paper. There exist different types of copulas and the most common used families are elliptical and Archimedean copulas. Among the elliptical copulas we will use the Gaussian and Student-t copula and among the Archimedean family we will focus on the Clayton copula.

Elliptical copulas
Elliptical copulas, such as the Gaussian and Student-t, are called implicit copulas since they are constructed by the inversion method instead of an explicit formula. The elliptical copulas, given by function (4), are symmetrical which means that they have the same strength of dependence in both the upper and lower tails.

Gaussian copula (normal)
The Gaussian copula function is the most common implicit copula used in finance. The main reason for its popularity is that the Gaussian copula is very easy to implement since the parameters for calibration are given by the correlation matrix, which we now denote $\Sigma$. The main weakness with the Gaussian copula is its symmetry which means that it can not handle strong and asymmetric tail dependence. Let $\Phi$ and $\phi$ be the multivariate and the univariate standard normal distribution functions respectively. Then the Gaussian copula is defined by:

$$C(u_1, \ldots, u_d; \Sigma) = \Phi(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)).$$

Unfortunately the Gaussian copula can only be expressed in integral form and it is therefore easier to work with the Gaussian copula density function. See McNeil et al. (2005, 191) for the mathematical expression of the integral form in two dimensions.

Further, if we let $\xi = (\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d))$, then the Gaussian copula density function is given by differentiating (6) (Bouyé et al., 2000, 16; Alexander, 2008b, 268):

$$c(u_1, \ldots, u_d; \Sigma) = |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} \xi^T(\Sigma^{-1} - I) \xi \right),$$

where $|\Sigma|$ is the determinant of the correlation matrix.

Student-t copulas
The Student-t copula is another implicit elliptical copula. As we can see in Figure 10, the Student-t copula is often symmetric (it is possible to construct an asymmetric Student-t copula) like the Gaussian copula but the difference is that it has higher values in the tails. This implies that the Student-t copula is a better choice than the Gaussian copula if we want a copula that can handle stronger tail dependence. Though it comes with the cost of estimating another parameter, called degrees of freedom, $\nu$. Let $t_v$ and $t_v$ be the multivariate and univariate Student-t distribution function respectively, with $v$ degrees of freedom. Further we let $\Sigma$ be the correlation matrix. Then the Student-t copula is defined by:

$$C_v(u_1, \ldots, u_d; \Sigma) = t_v(t_v^{-1}(u_1), \ldots, t_v^{-1}(u_d)).$$

24
The Gaussian copula and the Student-t copula are related to each other such that when \( \nu \to \infty \), the Student-t copula will turn into a Gaussian copula. Like the Gaussian copula, the Student-t copula can not be written in a simple closed form. But by inserting the Student-t density function in (7) and differentiating, we obtain the corresponding Student-t copula density function (Bouyé et al., 2000, 16; Alexander, 2008b, 268):

\[
c_{\nu}(u_1, \ldots, u_d; \Sigma) = K|\Sigma|^{-1/2}(1 + \nu^{-1}\xi^T\Sigma^{-1}\xi)^{-(\nu+d)/2}\prod_{i=1}^{d}(1 + \nu^{-1}\xi_i^2)^{(\nu+1)/2},
\]

where \( \xi = (t_{\nu}^{-1}(u_1), \ldots, t_{\nu}^{-1}(u_d)) \) and \( K \) is given by the gamma function \( \Gamma \):

\[
K = \Gamma\left(\frac{\nu}{2}\right)^{-d/2} \Gamma\left(\frac{\nu+1}{2}\right)^{-d} \Gamma\left(\frac{\nu + d}{2}\right).
\] (8)

**Why turn to other copulas?**

As mentioned in section 2.2 about stylized facts, stock returns have been observed to have higher dependence for large negative returns compared to large positive returns. This implies that stock returns have an asymmetric tail dependence and means that the dependence is not constant for all market conditions. In contrast to the Gaussian and Student-t copulas there are a number of copulas which can be stated directly and have a quite simple form. These copulas, which typically belongs to the family of Archimedean copulas, are asymmetrical and only have tail dependence in one corner. They are therefore said to be adequate for financial applications.

**Archimedean copulas**

In contrast to implicit copulas, Archimedean copulas are often called explicit copulas since they can be constructed using an explicit formula. Within the Archimedean copulas this paper will focus on the one parameter Clayton copula. This means that there is only one parameter that form the distribution and determines the strength of dependence.

The Archimedean copulas are constructed by using a generator function \( \Psi(u) \) and its pseudo-inverse \( \Psi^{-1}(u) \). A continuous, strictly decreasing convex function \( \Psi : [0, 1] \to [0, \infty] \) which satisfies \( \Psi(1) = 0 \), is defined as a *Archimedean copula generator*. If \( \Psi(0) = \infty \), it is also known as a strict generator.

Let \( \Psi(u) \) be a Archimedean generator function that fulfills \( \Psi(0) \leq \infty \). Then the function \( \Psi \) has a *pseudo inverse*, \( \Psi^{-1} \) defined by (Alexander, 2008b, 271; McNeil et al., 220ff):

\[
\Psi^{-1}(u) = \begin{cases} 
\Psi^{-1}(u), & 0 \leq u \leq \Psi(0), \\
0, & \Psi(0) < u \leq \infty.
\end{cases}
\]

Given any generator function \( \Psi(u) \), an *Archimedean copula* is defined by:

\[
C(u_1, \ldots, u_d) = \Psi^{-1}(\Psi(u_1) + \ldots + \Psi(u_d)).
\]
The corresponding density function is given by:

\[ c(u_1, \ldots, u_d) = \Psi^{-1}_d(\Psi(u_1) + \ldots + \Psi(u_d)) \prod_{i=1}^{d} \Psi'(u_i), \]

where \( \Psi^{-1}_d(u) \) is the dth derivative of the inverse generator function and \( \Psi'(u) \) denotes the derivative of \( \Psi(u) \).

Archimedean copulas can be created with more than one parameter but it is most common to only use one. However, constructions of high dimensional one parameters Archimedean copulas can easily fail since they are only limited to one parameter. A necessary condition when using d-dimensions is that the generator function \( \Psi(u) \) is strict, if the inverse of the generator function \( \Psi^{-1}(u) \) is completely monotonic. Although one parameter Archimedean copulas often are criticized by researchers because of this limitation, they are frequently used in practice, see Embrechts and Hofert (2011) for a discussion.

**Clayton copula**

The Clayton copula is one of several Archimedean copulas and was first studied by Clayton (1978). This type, which can be seen in Figure 10 above, has become popular in financial risk modelling since it can handle asymmetric tail dependence and has a strong dependence in the lower tail. The Clayton copula is described by the parameter \( \alpha \) and can be constructed in \( d \) dimensions since the generator function is strict and its inverse is completely monotonic. Given the generator function:

\[ \Psi(u) = \alpha^{-1}(u^{-\alpha} - 1), \quad \alpha \neq 0, \]

and the corresponding inverse generator:

\[ \Psi^{-1}(x) = (\alpha x + 1)^{-1/\alpha}, \]

the Clayton copula is defined by:

\[ C(u_1, \ldots, u_d; \alpha) = (u_1^{-\alpha} + \ldots + u_d^{-\alpha} - d + 1)^{-1/\alpha}. \]  \hspace{1cm} (9)

By differentiating (9) we obtain the Clayton copula density function:

\[ c(u_1, \ldots, u_d; \alpha) = \left(1 - d + \sum_{i=1}^{d} u_i^{-\alpha}\right)^{-d-(1/\alpha)} \prod_{j=1}^{d} (u_j^{-\alpha}(j - 1)\alpha + 1). \]  \hspace{1cm} (10)

As the parameter \( \alpha \) varies, the Clayton copulas capture a range of dependence and perfect positive dependence occurs when \( \alpha \to \infty \).

**Calibration of copulas**

Like any other multivariate joint distribution, a copula is described by parameters that determines its distribution. This section describes how to estimate these parameters, denoted by \( \theta \), of the parametric copula \( C_{\theta} \). Calibration of copulas has been widely studied

There are two main methods to calibrate copula parameters are called the Method-of-Moments and the maximum likelihood method. In most cases we need use the maximum likelihood method because the Method-of-Moments can only be used for a limited amount of copulas even though it is often easier to use.

**Method-of-Moments using rank correlation**

With this method we look for a theoretical relationship between one of the rank correlations and the parameters of the copula. The empirical values of the rank correlation are substituted into this relationship to get an estimate of the copula parameters. The advantage with this method is that we do not need to estimate the marginals. Both Nelsen (2006, 168) and Alexander (2008b, 280) presents the following relationship between a bivariate copula $C(u,v)$, Kendall’s tau, $\tau$, and Spearman’s rho, $\rho_S$:

\[
\tau = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1, \quad (11)
\]

\[
\rho_S = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3. \quad (12)
\]

By inserting the Gaussian or Student-t copula in (11) or (12) and solving for the Pearson’s correlation $\rho$ we get:

\[
\rho = 2\sin\left(\frac{\pi}{6}\rho_S\right),
\]

and

\[
\rho = \sin\left(\frac{\pi}{2}\tau\right).
\]

For the Archimedean copulas we can express Kendalls tau as:

\[
\tau = 1 + 4 \int_0^1 \left(\frac{\Psi(x)}{\Psi'(x)}\right) dx, \quad (13)
\]

where $\Psi'(x)$ is the first derivative of the generator function $\Psi(x)$. When applying (13) to a Clayton copula we obtain:

\[
\tau = \frac{\alpha}{\alpha + 2} \Rightarrow \alpha = \frac{2\tau}{(1 - \tau)}. \quad (14)
\]

However, the estimated correlation matrix, which often is called the pseudo-correlation may not necessarily be positive-semi definite although it very often is. This should always be tested and the matrix has to be corrected if needed by for example using the spectral decomposition method developed by Rebonato and Jäckel (1999). We refer the reader to
McNeil et al (2005, 231) for this algorithm.

**Maximum likelihood method (MLE)**

We now turn to the method called maximum likelihood (MLE), which is another method that can be used to estimate copula parameters. With this method we construct the copula density function to be as close as possible to the empirical copula density.

Suppose we have a joint density function of the form (5), where the marginal densities and distributions are given by $f_i(x_i; \alpha)$ and $F_i(x_i; \alpha)$ respectively for $i = 1, \ldots, d$. For simplicity we also let $\alpha$ be the single parameter for each marginal. The MLE is then obtained by rewriting (5) in the form of the following log likelihood function which then is maximized:

$$
\ln L(\alpha, \theta, x_1, \ldots, x_T) = \sum_{t=1}^{T} \ln c(F_1(x_{1t}; \hat{\alpha}_1), \ldots, F_d(x_{dt}; \hat{\alpha}_d); \theta) + \sum_{i=1}^{d} \sum_{t=1}^{T} \ln f_i(x_{it}; \alpha_i),
$$

where $\alpha = (\alpha_1, \ldots, \alpha_d)$ is the vector of marginal parameters, $\theta$ is the vector of copula parameters, $x_t = (x_{1t}, \ldots, x_{dt})$ is the vector of observations on the $d$ r.vs. at time $t$. The MLE can be conducted in one single optimization but it is more time efficient to split it into two parts so that we first estimate the marginals and then calibrate the copula parameters (Alexander, 2008b, 281-282):

1. Calibrate the parameters for each marginal density:
   $$\max_{\alpha_i} \sum_{t=1}^{T} \ln f_i(x_{it}; \alpha_i), \quad \text{for } i = 1, \ldots, d.$$

2. Calibrate the copula parameters:
   $$\max_{\theta} \sum_{t=1}^{T} \ln c(F_1(x_{1t}; \hat{\alpha}_1), \ldots, F_d(x_{dt}; \hat{\alpha}_d); \theta).$$

There exist two methods under this approach. These are called inference on margins (IFM) and canonical maximum likelihood (CML). Which one to use depends on how we want to estimate the marginals. McNeil et al. (2005, 232-233) lists three possible methods for for estimating the marginal distributions. When using the parametric approach for estimating the marginals the method is called IFM. In the opposite, when estimating the marginals based on the empirical distributions the calibration method is often called canonical maximum likelihood (CML). Since we will estimate our marginals by a method called Extreme Value Theory for tails (EVT), this implies that we use a CML method to optimize our copula parameters for the Gaussian and Student-t copula. For the Clayton copula parameter we will use a version of the Method-of-Moments which will be explained in more detail below.

**Calibrating Archimedean copula**

Hofert et al. (2012) have presented an excellent review many different ways of estimate Archimedean copula parameters, such as Method-of-Moments, minimum distance estimation and MLE. In this paper we will use the Method-of-Moments using pairwise samples of Kendall’s tau. Consequently we will calculate Kendalls’s sample coefficient for all pairwise
asset returns and then by taking the arithmetic mean of theses coefficient, use relationship (14) to estimate the parameter $\alpha$. Although, according to Hofert et al. (2012) there is no theoretical justification for applying this method in dimensions $d > 2$, it is a simple method which has been used by Berg (2009) and Savu and Trede (2010).

**Goodness of fit**

Goodness-of-fit (GoF) is a method to determine if the chosen copula fits the historical data. Both Alexander (2008b), Genest and Favre (2007) presents an overview of available GoF techniques that can be used. When there are few parameters for the copula, Alexander (2008b, 283) prefers the straightforward method to compare the values of the optimized likelihood function. However, in this paper we will use the mean square error (MSE) to determine if the chosen parametric copula fits the empirical copula. MSE is calculated by the square root of the sum of the squared differences between the empirical copula distribution and the fitted copula. A good fit between the chosen copula and the data is given by a small value of MSE. Let the empirical copula, denoted by $C^E$, be given by:

$$C^E(v_1, \ldots, v_d) = \frac{1}{d} \sum_{t=1}^{d} \mathbb{1} \{u_{t,1} \leq v_1, \ldots, u_{t,d} \leq v_d\},$$  

(15)

where $\mathbb{1}_A(a)$ is the indicator function of a given set $A$, i.e. $\mathbb{1}_A(a) = 0$ if $a \notin A$ and $\mathbb{1}_A(a) = 1$ if $a \in A$. Then the MSE is defined by:

$$MSE = \frac{1}{d} \sum_{t=1}^{d} (C^E_t - C_t)^2,$$

where $d$ is the number of data points, $C$ is the fitted copula and $C^E$ is the empirical copula given by (15).

**Simulation of copulas**

As we have seen above, when calibrating copulas the procedure is first to calibrate the marginals and then the copula parameters. In simulation it is the other way around. First we simulate the dependence and then the marginals. The fact that we can apply arbitrary copulas to arbitrary marginals is an important concept with copulas. Alexander (2008b) puts it this way: “And the wonderful thing about copulas is that the distributions of the marginals can all be different, and different from the copula.”. For example, the marginals do not need to be normally distributed to be modelled by a Gaussian copula.

There are several ways to simulate from copulas but the three most common ones are; simulation with conditional copulas, simulation with elliptical copulas and the Marshall and Olkin algorithm. Below we will present the latter two, which also are the ones used in this paper. An example of how simulations from the Gaussian, Student-t and Clayton copula is presented in Figure 11.
Simulation from elliptical copulas

Before presenting the algorithm we need to explain the term Cholesky matrix. Given a positive definite matrix $A$, a Cholesky matrix $L$, is a lower triangular matrix such that:

$$A = LL^T,$$

and has the same dimensions as $A$. Matrix $A$ is called a lower triangular matrix if all the values above the main diagonal is equal to zero. The Cholesky matrix is obtained by Cholesky decomposition and we refer the reader to Alexander (2008a, 62) for more details about Cholesky decomposition.

When simulating from elliptical copulas we assume that the joint distribution we want to simulate from is given by:

$$F_{X_i}(x_i) = \begin{cases} 
\Phi(x_i), & \text{for the Gaussian distribution}, \\
t_{\nu}(x_i), & \text{for the Student’s-t distribution}
\end{cases}$$

To obtain simulated returns we then use the Cholesky matrix of the covariance matrix by the following algorithm.
**Algorithm 2:** Simulations from elliptical copulas

1. Generate simulations $u_1, \ldots, u_d$, from independent uniformed r.v.s.
2. Set $x_i = F_{X_i}^{-1}(u_i)$ and apply the Cholesky matrix of the covariance matrix on \{ $x_1, \ldots, x_d$ \} to obtain simulations \{ $x_1^*, \ldots, x_d^*$ \}.
3. Set $u_i^* = F_{X_i}(x_i^*)$ to obtain a simulation \{ $u_1^*, \ldots, u_d^*$ \}.
4. Use the inverse marginal distributions to obtain $x_1, \ldots, x_d = (F_{X_1}^{-1}(u_1^*), \ldots, F_{X_d}^{-1}(u_d^*))$.

**Simulation of Archimedean copulas**

In the following algorithm we describe in general how to simulate from Archimedean copulas and in particular from a Clayton copula. When simulating from high-dimensional Archimedean copulas the Marshall and Olkin algorithm is popular to use. This algorithm is based on Laplace–Stieltjes transformation, an integral transformation technique, which is used to construct random vectors whose distributions are multivariate Archimedean copulas. See McNeil et al. (2005, 223) for more details.

**Algorithm 3:** Marshall and Olkin algorithm

1. Generate a variate $V$ with distribution function $G$, where $\hat{G}$, the Laplace–Stieltjes transform of $G$, is the inverse of the generator $\Psi$ of the required copula.
   
   - For simulations from the Clayton copula we generate a variate from the gamma distribution: $V \sim Ga(1/\theta, 1)$ with $\theta > 0$. The distribution function of $V$ has the Laplace transform $\hat{G}(t) = (1 + t)^{-1/\theta}$.

2. Generate variates $u_1, \ldots, u_d$, from an independent uniform distribution.
3. Return $u_1^*, \ldots, u_d^* = (\hat{G}(-ln(u_1)/V), \ldots, \hat{G}(-ln(u_d)/V))^T$.
4. Use the inverse marginal distribution to obtain the returns $x_1, \ldots, x_d = F_{X_1}^{-1}(u_1^*), \ldots, F_{X_d}^{-1}(u_d^*)$. 
2.4 Risk metric

Portfolio optimization is a method for choosing a certain allocation of assets based on estimates of the expected return and the market risk. A risk metric states the magnitude of the uncertainty of the expected return and can be measured in different ways. In this section we will explain our choice of risk metric. From a mathematical point of view, a risk metric is a function that maps r.vs. to the real numbers (Alexander, 2008d). Consider a given probability space \((\Omega, \mathcal{A}, \mathbb{P})\) and let \(V\) be a non-empty set of \(\mathcal{A}\)-measurable real-valued r.v.s., then any mapping \(\varphi : V \Rightarrow \mathbb{R} \cup \{\infty\}\) is called a risk metric.

Artzner et al. (1999) have proposed a number of properties that such a risk metric should fulfill. If the risk metric satisfies these properties it is called coherent. A function \(\varphi(\cdot) : L^\infty \Rightarrow \mathbb{R}\) is called a coherent risk metric if it for \(X, Y\) and a constant \(c\), satisfies the following properties:

1. Monotonicity
   \[X, Y \in V, \ X \leq Y \Rightarrow \varphi(X) \geq \varphi(Y).\]

2. Subadditivity
   \[X, Y, X + Y \in V \Rightarrow \varphi(X + Y) \leq \varphi(X) + \varphi(Y).\]

3. Positive homogeneity
   \[X \in V, \ c > 0 \Rightarrow \varphi(cX) = c\varphi(X).\]

4. Translation invariance
   \[X \in V, \ c \in \mathbb{R} \Rightarrow \varphi(X + c) = \varphi(X) - c.\]

Volatility

According to Danielsson (2011, 9) the most widely used risk metric is the volatility, which is defined as the standard deviation of returns. As mentioned in previous sections, volatility assumes normality which could lead to inaccurate conclusions about the riskiness in the assets.

Value at Risk

An alternative risk metric in order to determine the risk in financial returns is Value at Risk (VaR). This risk metric estimates the potential loss based on a time period and a given probability level. A widely accepted definition of VaR is as follows: “Value at risk is the maximum amount of money that may be lost on a portfolio over a given period of time, with given level of confidence”. In mathematical terms, let \(\alpha \in (0, 1]\) be fixed and \(X\) be a real-valued r.v. on the probability space \((\Omega, \mathcal{A}, P)\) and define \(\inf \emptyset = \infty\). The \(\alpha\)-quantile of \(X\) is then:

\[q_\alpha(X) = \inf \{x \in \mathbb{R} : P[X \leq x] \geq \alpha\},\]
Theory

and the VaR at confidence level $\alpha$ of $X$ is given by:

$$VaR_\alpha(X) = q_\alpha(-X).$$

Alexander (2008d) stated several attractive features that VaR has captured as a risk metric compared to the volatility. However, this risk metric has provoked a wide debate among quants and academics since it is not necessarily is sub-additive which is not in line with the properties of a coherent risk metric.

Conditional Value at Risk

A risk metric that satisfies all the properties of a coherent risk metric is the Conditional Value at Risk (CVaR). This metric summarizes the entire tail of the distribution as a single risk measurement and estimates the expected loss given that the loss exceed VaR. Let $\alpha \in (0, 1]$ be fixed and $X$ be a real-valued r.v. on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $E[\max(0, -X)] < \infty$. The CVaR at confidence level $\alpha$ of $X$ is then:

$$ES_\alpha(X) = -(1 - \alpha)^{-1} \left( E[X] \mathbb{1}_{\{-X \geq q_\alpha(-X)\}} + q_\alpha(-X) \left\{ \alpha - \mathbb{P}[-X < q_\alpha(-X)] \right\} \right),$$

where $\mathbb{1}_A(a)$ is the indicator function of a given set $A$, i.e. $\mathbb{1}_A(a) = 0$ if $a \notin A$ and $\mathbb{1}_A(a) = 1$ if $a \in A$. By using proposition 3.2 in Acerbi and Tasche (2002), we can formulate CVaR in terms of VaR as:

$$ES_\alpha = (1 - \alpha)^{-1} \int_\alpha^1 VaR_u(X) \, du.$$  \hfill (16)
2.5 Estimation of the marginal distributions

The portfolio optimization in this paper is based on simulated returns based on copula functions which describes the dependence structure. Up till now we have stated how to simulate returns and what properties an adequate risk metric should fulfill. But before we are able to simulate returns, we need to prepare for copula modelling by characterizing the distribution of returns of each asset. Hence, in this section we focus on how the to model the cumulative distribution function for each asset before calibrating the copula. There are three main methods that can be used for estimating the marginal distributions; the non-parametric, the semi-parametric and the parametric.

Parametric and non-parametric methods

The non-parametric method refers to historical simulation (HS), which determines the risk only based on the empirical distribution of the returns. This method relies on the assumption that the historical movements of the asset returns will repeats itself. Danielsson (2011, 95-97) states several strengths with this method compared to the parametric-based methods. HS is less sensitive to extreme values and does not incorporate estimation error in the same way as the other methods. Furthermore, when we use HS we do not make any assumptions about the distribution of returns. This implies that HS can capture complex dependence structures in a way that parametric-based methods can not. However, there exist several weaknesses by using this approach. The sample size has to be relatively large compared to parametric methods which implies that it will take longer time for the forecast to adjust to changes. Also, each historical observation has the same impact in the forecast which could be a disadvantage, particularly when there has been a change in volatility. The parametric method, also known as the "Variance-covariance"-method, makes an assumption that the future returns could be described by some distributional assumptions as normality or Student-t. However, if the returns do not follow the chosen distribution, the computed market risk will be misleading and could underestimate the actual risk in the financial asset.

Semi-parametric method

The semi-parametric approach can be described as different methods that combines the parametric and the non-parametric approaches. There exist several different models in this area but this paper will use Extreme Value Theory (EVT). Most of the fully statistical methods focus on modelling the whole distribution of the returns, where most of the observations are located in the centre of the distribution. Accordingly, these methods will approximate common events very accurate but will give misleading conclusions about the distributions of the tails. Compared to these methods, EVT focuses explicitly on analyzing the tail of the distribution. By doing this, this method will lead to more accurate estimations for extreme events. Roughly speaking, there are two main methods that models extreme values: Block maxima or Peak Over Threshold (POT). This paper will focus on the latter since it is generally considered to be more efficient then the Block maxima method (McNeil et al., 2005, 275). There exist two approaches to model POT: The generalized Pareto distribution (GPD) and the Hill Estimator (HE). In this paper
we will use GPD to estimate CVaR. The following derivation of CVaR based on GPD is based on the presentation of Extreme Value Theory by McNeil et al. (2005). This paper is recommended for the reader whenever theoretical details are missing.

In order to define the generalized Pareto distribution we first have to define the conditional excess distribution function $F_u$. Let $X$ be a r.v. with distribution function $F$. Let the threshold be denoted by $u$ and let $y = x - u$ be the excesses. For example, given a distribution of daily returns, then $u = 0.02\%$ implies that we are interested to model the observations that exceeds $0.02\%$. In practice we will choose the threshold such that we fit the Pareto tails to the distribution at a certain cumulative probabilities, for example 0.1 and 0.9. The distribution $F_u(y)$ is given by:

$$F_u(y) = P(X - u \leq y | X > u) = \frac{F(x) - F(u)}{1 - F(u)}$$  \hspace{1cm} (17)

for $0 \leq y \leq x_F - u$ where $x_F \leq \infty$ is the right endpoint of $F$. The estimation of the function $F$ is quite straightforward since we normally have a lot of observation in this area. However, it can be much more complicated to estimate the distribution of the tails since there in general are less observations in this area. In order to overcome this problem, Pickands, Balkema and de Haan (1974) came up with a powerful theorem which will help us approximate the conditional excess distribution function $F_u(y)$.

**Pickands–Balkema–de Haan theorem.** For a large class of underlying distribution functions $F$, and large $u$, $F_u(y)$ is well approximated by the generalized Pareto distribution (GPD):

$$F_u(y) \approx G_{\varepsilon, \beta}(y), \quad u \to \infty,$$

where

$$G_{\varepsilon, \beta}(y) = \begin{cases} 1 - \left(1 + \frac{y}{\beta}\right)^{-\frac{1}{\varepsilon}}, & \varepsilon \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta}\right), & \varepsilon = 0 \end{cases}$$  \hspace{1cm} (18)

for $0 \leq y \leq x_F - u$, is the so-called generalized Pareto distribution (GPD). The parameters $\beta$ and $\varepsilon$ is the *scale parameter* and *shape parameter*, respectively. By defining $x = u + y$, the GPD can be formulated as a function of $x$:

$$G_{\varepsilon, \beta}(x) = 1 - \left(1 + \varepsilon \frac{x - u}{\beta}\right)^{-\frac{1}{\varepsilon}}.$$  \hspace{1cm} (19)

In order to find expressions for the VaR and CVaR as a function of GPD parameters, we recall (17) for the conditional excess distribution and solve it for $F(x)$

$$F(x) = (1 - F(u))F_u(x) + F(u)$$

By substituting the $F_u$ with the GPD and then estimating $F(u)$ with the simple empirical estimator:

$$F_u = \frac{n - N_u}{n},$$
where \( n \) is the number of observations and \( N_u \) is the number of observations above the threshold \( u \), we obtain the following estimation of \( F(x) \):

\[
F(x) = \frac{N_u}{n} \left( 1 - \left( 1 + \frac{\varepsilon}{\beta} (x - u) \right)^{-1/\varepsilon} \right) + \left( 1 - \frac{N_u}{n} \right) = 1 - \frac{N_u}{n} \left( 1 + \frac{\varepsilon}{\beta} (x - u) \right)^{-1/\varepsilon}. \tag{20}
\]

By inverting (20) for a given confidence level \( \alpha \), we obtain the expression for \( \text{VaR}_\alpha \):

\[
\text{VaR}_\alpha = u + \frac{\beta}{\varepsilon} \left( \left( \frac{n}{N_u} \alpha \right)^{-\varepsilon} - 1 \right).
\]

Furthermore, assuming that \( \varepsilon < 1 \), the CVaR can be estimated as:

\[
\text{CVaR}_\alpha = \frac{\text{VaR}_\alpha}{1 - \varepsilon} + \frac{\beta - \varepsilon u}{1 - \varepsilon}.
\]

The parameters \( \varepsilon \) and \( \beta \) can be estimated from the log-likelihood function of the GPD. Let \( g_{\varepsilon, \beta} \) be the density function of a given GPD, the log-likelihood is then:

\[
\ln L(\varepsilon, \beta; Y_1, \ldots, Y_{N_u}) = \sum_{j=1}^{N_u} \ln g_{\varepsilon, \beta}(Y_j) = -N_u \ln \beta - \left( 1 + \frac{1}{\varepsilon} \right) \sum_{j=1}^{N_u} \ln \left( 1 + \varepsilon \frac{Y_j}{\beta} \right).
\]

How many observations that are needed in the tail is equivalent to setting the choice of the threshold \( u \) to a suitable value. According to Embrechts et al. (2003, 11-12) there is no easy answer to this question and it all depends on the amount of data that are used. However, a minimum of 25 exceedances seems to be an accurate guideline when choosing the threshold.

**Kernel smoothing**

Kernel smoothing is a commonly used non-parametric method in order to represent the pdf or cdf of a r.v.. In this paper we will use a Kernel smoothing method for the interior of the distribution and then use EVT for the tails of the distribution. By using Kernel density estimation we will make inferences about the distribution based on a finite data sample. This method is good to use in situations where a parametric distribution cannot describe the r.v. or when one want to avoid making assumptions about the distribution.

If we consider a i.i.d. sample drawn \((x_1, \ldots, x_n)\) from some unknown distribution \( f \). The **Kernel estimator** is then formulated as:

\[
\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right),
\]

where \( n \) is the sample size, \( K(\cdot) \) is the Kernel smoothing function and \( h \) is a parameter called the bandwidth. Since Kernel smoothing method is not a focus area of this paper, we do not go into further details. Instead we refer the reader to Wand and Jones (1995) for a more detailed description about the method and how to choose the parameters.
Application in practice
In Figure 12 we can see both the empirical observations (blue marks) and the fitted cdf (red line). The cdf is created by using the Pareto tail distribution for the lower and upper tail. Further, the interior between the tails is created by using Kernel smoothing.

Figure 12: Empirical observations and the fitted cdf created by Pareto tail distribution and Kernel smoothing.
2.6 Portfolio optimization

The classic portfolio allocation problem is often stated: How should a rational investor allocate the funds between a set of different risky assets? This leads us to the portfolio optimization problem where we want to minimize the portfolio risk with respect to a set of constraints such as portfolio weights and a level of expected return. In this section we start by explaining the classic mean-variance optimization created by Markowitz (1952) and then move on to present another portfolio optimization approach using CVaR created by Rockafeller and Uryasev (1999).

Optimization in normal markets

Harry Markowitz revolutionized the modern finance theory with modern portfolio theory (MPT). His approach is called mean-variance portfolio optimization and means that we can select a portfolio that minimizes the risk with respect to the expected return or in the opposite maximizes the expected return given a level of risk. A lot of researchers have been inspired by Markowitz paper and developed several models based on his work. Some of the most famous models are the Capital Asset Pricing Model (CAPM) developed by William Sharpe and the Arbitrage Pricing Theory (APT) developed by Stephen Ross. However, Markowitz theory has been criticized by several academics for being inadequacy and misleading since this approach assumes normally distributed returns which implies that it can not capture market asymmetries and extreme events. However, in 1999 Rockafeller and Uryasev presented a new approach of how to take these real world financial time series behaviour into consideration when optimizing a portfolio.

Mean-Variance optimization

The mathematical representation of this problem can vary depending on the objective function the investor would like to optimize. This paper will define a model which will generate a portfolio with the lowest expected risk level, given a required minimum expected return. Let \( n \) be the number of instruments and \( x \) be a decision vector that represents a portfolio of financial instruments where \( X = \{ x_1, \ldots, x_n \} \). Further, let \( S \) denote the feasible set of portfolios, which could be expressed as:

\[
S = \left\{ X \in \mathbb{R}^n : x_i \geq 0 \ \forall \ i \in \{1, \ldots, n\}, \sum_{i=1}^{n} x_i = 1 \right\} \tag{21}
\]

Further, let \( y = \{ y_1, \ldots, y_n \} \) be a random vector denoting the return on instrument \( i \). The loss of a portfolio \( X \) expressed as a function of \( X \) and \( y \) is therefore given by:

\[
f(X, y) = -[x_1 y_1 + \cdots + x_n y_n] = -X^T y.
\]

Finally, we have to find an expression for the required expected return \( R \). The investors minimum required return constraint could be formulated as

\[
X^T y \leq -R
\]

Having defined all the variables and parameters, we can now can define the minimum variance portfolio. Let \( \Sigma \) be the covariance matrix of \( y \) and \( S \) be a set of possible portfolios...
given by (21). The minimum variance portfolio can then be obtained by solving:

$$\begin{align*}
\min & \quad X^T \Sigma X \\
\text{subject to} & \quad X^T y \leq -R \\
& \quad X \in S
\end{align*}$$

Efficient frontier
The efficient frontier is a term introduced by Markowitz (1952) which refers to a set of optimal portfolios in a risk-return diagram. Let the expected return of the assets be denoted by $\mu = E(y)$ and the portfolio standard deviation be denoted by $\sigma = (X^T \Sigma X)^{1/2}$. The set of all points $\{(X^T \Sigma X)^{1/2}, X^T y\}$ is called the feasible set in the $\{\mu, \sigma\}$-space and can be illustrated using a mean-variance diagram. One can also change the variance to CVaR and illustrate the efficient frontier in a mean-CVaR diagram as in our case. In Figure 13 we present six stock market indices that each represents an expected return and a risk level given by CVaR. The market indices, which will be presented in a coming section, can be combined into portfolios in different combinations. The bold blue line is called the efficient frontier or "Markowitz bullet" and represents a set of optimal portfolios. Given an efficient portfolio, no other portfolio exists with a higher expected return given the same level of risk. In the opposite way we can neither construct a portfolio with a lower risk, given the same expected return (Alexander, 2008b, p. 246ff). Since we use CVaR instead of variance it is important to know that if the return-loss distribution is normal then the efficient frontier will be the same for both CVaR and variance (Rockafeller and Uryasev, 2000).

There exist several ways to select the optimal portfolio. Some of the most common portfolios are the minimum-variance portfolio and the Sharpe portfolio. However, in this paper we focus on the minimum CVaR portfolio (also called min CVaR) and portfolios based on different levels of expected return, $r$. We refer the reader to Alexander (2008b) for a more detailed description of other optimal portfolios.

In Figure 13 we present examples of two optimized portfolios. The red circle is a minimum CVaR portfolio with the constraint $r \geq 0.05\%$ and the green circle is an optimized minimum CVaR portfolio with no constraint of expected daily return. It is important to note that when the required level of expected return is outside the efficient frontier, then the optimized portfolio will either be the minimum CVaR portfolio or the portfolio with the maximum expected return. For example, if we want to minimize the CVaR given an expected return of minimum 0.02 %, the optimal portfolio in Figure 13 will be the green circle.
Optimization in non-normal markets
This approach is based on the presentation of optimizing CVaR by Rockafeller and Uryasev (1999). The theoretical description of this approach is largely based on their result and whenever any details are missing the reader is recommended to their work.

CVaR optimization
As in the previous section, let \( y \) be a random vector and \( X \) denote a vector representing a given portfolio where \( S \) represents all possible portfolios. The probability that \( f(X, y) \) does not exceed a given threshold \( c \) is then:

\[
\Psi(X, c) = \int_{f(X,y)\leq c} p(y) \, dy
\]

Consequently, the VaR and the CVaR for some given confidence level \( \alpha \), denoted \( c_\alpha(X) \) and \( \phi_\alpha(X) \) respectively, are given by:

\[
c_\alpha(X) = \min\{c \in \mathbb{R} : \Psi(X, c) \geq \alpha\}
\]

\[
\phi_\alpha(X) = \frac{1}{1 - \alpha} \int_{f(X,y)\geq c_\alpha(X)} f(X,y) \, p(y) \, dy.
\]

The expression for CVaR in (23) is complicated to compute since it depends on VaR, \( c_\alpha(X) \). In order to overcome this problem, Rockafeller and Uryasev (1999) instead uses the derivative of (23), which is independent from equation (22). By doing this, the complexity of the optimization problem is reduced and the function is then given by:

\[
F_\alpha(X, c) = c + \frac{1}{1 - \alpha} \int_{y} [f(X,y) - c]^+ \, p(y) \, dy.
\]
Theory

According to their second theorem, minimizing (23) is equivalent to minimizing (24), i.e:

$$\min_{X \in S} \phi_\alpha(X) = \min_{(X,c) \in X \times \mathbb{R}} F_\alpha(X,c)$$

Furthermore, the expression for $F_\alpha(X,c)$ can be approximated by

$$\hat{F}_\alpha(X,c) = c + \frac{1}{q(1-\alpha)} \sum_{k=1}^{q} [f(X,y_k) - c]^+$$

which makes the expression $\hat{F}_\alpha(X,c)$ convex and piecewise linear.

Application in practice

Let $q$ be the number of i.i.d. samples from the loss vector $y$. A sample set $y = \{y_1, \ldots, y_q\}$ yields the following approximate function of $F_\alpha$:

$$\hat{F}_\alpha(X,c) = c + \frac{1}{q(1-\alpha)} \sum_{k=1}^{q} [-X^T y_k - c]^+$$

This problem can be reduced to convex programming in terms of auxiliary real variables. According to Collins et al. (2001), auxiliary variables can be explained as variables which can help to make estimates on incomplete data. Let $u_k$ be a set of auxiliary variables for $k = \{1, \ldots, q\}$, then the linear form of the optimization problem is given by:

$$\min \quad c + \frac{1}{q(1-\alpha)} \sum_{k=1}^{q} u_k$$

subject to

$$u_k \geq -X^T y_k - c$$

$$u_k \geq 0$$

$$X^T y \leq -R$$

$$X \in S$$

(25)
3 Method

In this section we describe the method that has been implemented for this study. First we present implementation and software before moving on to describe the data.

3.1 Implementation and software usage

The method we used in this paper can roughly be explained as a repeated portfolio optimization where the joint distributions were based on copula functions and the risk was measured with CVaR. The portfolios were optimized and backtested on monthly basis during 2000-2016. We invested by the optimal weights in six different indices with a buy-and-hold investment strategy for 22 trading days before the procedure was repeated by moving forward the calibration window. The optimization problems were solved in Matlab with the PortfolioCVaR object. Below we explain the step-by-step procedure in more detail:


As described in section 2.5, we used a semi-parametric method to model the marginal distributions. We adopted two different rolling calibration windows; the most recent 500 and 1250 observations of daily returns. They were chosen to approximately reflect two and five years of trading days. We are aware of the fact that this methodology leads to overlapping data samples which is not preferable if there are better options (Danielsson & Zhou, 2016). Though, this approach has been successfully and frequently used in portfolio optimization before, see for example Kakouris & Rustem (2014)

For the estimation of the marginals we used EVT for the tails and Kernel smoothing for the interior. The tails were modeled with generalized Pareto tails based upon the Peak over Threshold method. The threshold level was chosen such that 10 percent of the observations were placed in each tail. In other words, we fitted the Pareto tails to the distribution at the cumulative probabilities 0.1 and 0.9.

2. Quantile transformation of the returns using the inverse cdf.

At this step we transformed the observed daily returns of each index to uniform variables by quantile transformation. The theory about quantile transformation of random variables is described in section 2.1.2. At this step we created the empirical copulas in order to calibrate our chosen copulas in the next step.

3. Calibration of copula parameters.

Three different copula functions were calibrated: Gaussian, Student-t and Clayton, see section 2.3.3 for mathematical notations and illustrations. The Gaussian and Student-t copula parameters were calibrated with CML and the Clayton copula
was calibrated by Method-of-Moments. In particular, the Clayton parameter $\alpha$ was calculated by first estimating Kendall’s sample tau, see equation (2), for all pairwise combinations of indices. $\alpha$ was then given by converting each sample tau, $\tau$, to $\alpha$ by equation (14) and taking the arithmetic mean. This is also described in more detail in "Calibration of copulas" in section 2.3.3.

4. *Simulation of a large realization sample obtained from the estimated copula.*

When the copula parameters were estimated, we simulated jointly dependent uniform variables and then transformed them to daily returns via the inverse cdf of each index. The algorithms used for simulation are described under "Simulation of copulas" in section 2.3.3. For the Gaussian and Student-t copula we used Algorithm 2 and for the Clayton copula we used Algorithm 3. For each copula we ran 10 000 simulations, which according to Alexander (2008b) is an acceptable number of simulations for examination of the properties of copulas.

5. *Construction of the efficient frontier in terms of expected rate of return and CVaR.*

After we simulated returns for each asset, the risk was measured by CVaR, see equation (16) in section 2.4. The magnitude of CVaR is given by the confidence level $\alpha$. For this study we used $\alpha = 0.01$ since this is a common confidence level for computing market risk with CVaR. Furthermore, we calculated the expected return for each portfolio and constructed the efficient frontier as illustrated in Figure 13 in section 2.6.

6. *Selection of the optimal portfolio with respect to the objective function and the chosen constraints.*

In this step the optimal portfolios were chosen with respect to the objective function and the constraints. The portfolio constraints were given as follows:

- All problems were solved for different levels of expected daily return. The levels were set to $r = 0.01, 0.02$ and $0.03\%$. These daily returns corresponds to an annual return of approximate $2.5\%, 5.2\%$ and $7.9\%$ and were based on what we believe were reasonable expectations for the indices. For an example of an optimized portfolio based on a required expected daily return, see the red circle in Figure 13 in section 2.6. In addition we also chose to solve for the minimum CVaR portfolio, see the green circle in Figure 13 in section 2.6.
- The maximum weight for one single asset was set to $w_{\text{max}} \leq 40\%$. In a mathematical meaning, an upper bound in an optimization implies that we lower the feasible set of optimal portfolios. Roncalli (2011) showed that this modifies some properties of the empirical covariance matrix. Other than that, to the authors knowledge, there exists no guideline for how this constraint should be chosen. We chose this constraint since we believe it is a reasonable assumption for many investors.
Method

- The minimum weight for one single asset was set to $w_{min} \geq 0\%$. Thus, we only used non negative positions and in the same way as with the upper bound, we chose this constraint since we believe it is a reasonable and common way of investing.
- The sum of the weights equals one: $\sum_{i=1}^{d} w_i = 1$. Thus, all money had to be invested.

7. **Invest according to the optimal weights and apply a buy and hold strategy during a chosen number of trading days.**

After the portfolios had been selected, we adopted a buy-and-hold strategy for 22 trading days. The buy-and-hold investment strategy can shortly be explained as a passive investment strategy in which an investor buys assets and holds them for a period of time without reallocate the weights. We will not go into further details about this investment strategy since it is not the focus area in this paper. It is also important to note that we did not take into account transaction costs.

8. **Start over from step 1, moving the calibration window forward the number of chosen trading days.**

After we optimized the portfolios and invested by the optimal weights for 22 trading days, the process was repeated from step 1 by moving the calibration window forward.

3.2 Data

All data needed for this study was brought from Yahoo finance API for Matlab. We used the following six indices: CAC40, DAX, FTSE100, Nikkei225, S&P500 and TSX Composite. These indices corresponds to France, Germany, UK, Japan, US, and Canada respectively. The data used was daily returns for the period January 1995 – March 2016, see Appendix A for a visualisation of the chosen index. The chosen time period includes the important crises such as the dot-com bubble in 2000 and the financial crises in 2008. Below follows a short description of the chosen indicies:

- **CAC40 - France**
  CAC 40 is the most commonly used benchmark index for the French stock market. The index represents the performance of the 40 largest companies listed in France.

- **DAX - Germany**
  DAX is the largest and most commonly used benchmark index for the German stock market index. The index represent the performance of 30 selected German large companies traded on the Frankfurt Stock Exchange.

- **FTSE100 - UK**
  FTSE 100 index is an index which contains the 100 companies with highest market capitalization traded on the London Stock Exchange.
• **Nikkei225 - Japan**
  The Nikkei225 is a price-weighted average of 225 top rated Japanese companies listed at the Tokyo Stock Exchange.

• **S&P500 - US**
  Standard and Poor’s 500 Index is an American stock market index and represents 500 stocks traded from all major industries. This index is one of the most common index used in financial literature.

• **TSX Composite - Canada**
  The Toronto Stock Exchange Composite index represents over 75% of the Toronto Stock Exchange.
4 Results

The outline of this section is as follows. First we present the summary statistics so the reader can get a quick overview of the data. We then move on to present graphical results and finally we present results of different performance metrics for each optimized portfolio, sorted by the levels of expected return and the two length of calibration windows.

4.1 Summary statistics

In Table 1 and 2 we present daily and yearly statistical characteristics of the chosen indices. From these tables we can see that there is a difference in both mean and volatility between the indices. The summary statistics are based on the period 1995-2000 since the portfolios that were initially calibrated with five years of daily returns used data from 1995.

In Figure 14 we present the normalized return for each indices during the backtesting period 2000-2016. This graph shows that the backtesting period started in the beginning of the dot-com crises. All indices then rose until the financial crises in 2007/2008 and since then, we have observed a major recovery for a majority of the indices.

![Normalized index returns for the backtesting period 2000-2016.](image)

*Figure 14: Normalized index returns for the backtesting period 2000-2016.*
For the test of normality we divided the samples into 168 sub-periods, one for each period used for calibration, and conducted a Jarque-Bera test. The sub-periods consisted each of 1250 observations, thus we only tested the longer calibration window for normality. As we see from Table 1, the test rejected that the data came from a normal distribution for all sub-periods and indices. This fact is also partly supported by the QQ-plot, see Figure 27 in Appendix B where the entire sample distribution for all indices are illustrated. Consequently, this shows that the data had heavier tails compared to the normal distribution. Further, Figure 28 in Appendix B indicates that the pairwise correlation changed over time and for example rose significantly during the financial crises in 2007/2008. All together, these graphs and tests supported and justified the use of copulas, see section 2.2 for explanation.

Table 1: Statistical characteristics of the sample data from 1995 - 2016.

<table>
<thead>
<tr>
<th>Index</th>
<th>Daily mean return (%)</th>
<th>Daily volatility (%)</th>
<th>Jarque-Bera test</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSX Composite</td>
<td>0.0194</td>
<td>1.077</td>
<td>168</td>
</tr>
<tr>
<td>CAC40</td>
<td>0.0070</td>
<td>1.450</td>
<td>168</td>
</tr>
<tr>
<td>DAX</td>
<td>0.0254</td>
<td>1.485</td>
<td>168</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>-0.0209</td>
<td>1.497</td>
<td>168</td>
</tr>
<tr>
<td>FTSE100</td>
<td>0.0129</td>
<td>1.158</td>
<td>168</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>0.0261</td>
<td>1.202</td>
<td>168</td>
</tr>
</tbody>
</table>

Table 2: Mean, volatility and covariance for the yearly assets returns between 1995 and 2016.

<table>
<thead>
<tr>
<th>Index</th>
<th>Mean return</th>
<th>Volatility</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSX Comp.</td>
<td>0.0686</td>
<td>0.1548</td>
<td></td>
</tr>
<tr>
<td>CAC40</td>
<td>0.0738</td>
<td>0.2140</td>
<td>0.0237</td>
</tr>
<tr>
<td>DAX</td>
<td>0.1085</td>
<td>0.2328</td>
<td>0.0272</td>
</tr>
<tr>
<td>Nikkei225</td>
<td>0.0369</td>
<td>0.2503</td>
<td>0.0237</td>
</tr>
<tr>
<td>FTSE100</td>
<td>0.0355</td>
<td>0.1454</td>
<td>0.0285</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>0.0895</td>
<td>0.1744</td>
<td>0.0191</td>
</tr>
</tbody>
</table>

TSX Comp.  CAC40  DAX  Nikkei225  FTSE100
4.2 Backtesting results

In this section we provide graphical results from the backtesting procedure. The portfolios that were optimized and backtested are denoted: Normal, Gaussian, Student-t, Clayton and Equal. In the tables and graphs, Markowitz portfolio is denoted ”Normal” since the portfolio is based on normally distributed returns. The portfolio denoted ”Equal” is an equally weighted portfolio which was not optimized and chosen as a benchmark. We want to emphasize that the results are evaluated and discussed based on the portfolios performance during the entire period and not only based on a fixed end date.

Minimum CVaR

In this part we present the results from the minimum CVaR portfolios, exemplified by the green circle in Figure 13. Hence, for this portfolio we did not state any requirement of expected return. From Figure 15 we see that it was almost an equal performance for all portfolios until 2012 when they started to spread apart from each other. The same can be seen in Figure 16 where we used a rolling window of 1250 trading days for calibration. Even though the portfolios started to spread apart slightly earlier than 2012 in this figure, this states that it is of great importance not to focus on one fixed end date since changing end date to 2012 would have given a totally different result for the total return.

By visual inspection, we see that Markowitz portfolio based on normally distributed returns performed well with both calibration windows. Both graphs also indicates that the equally weighted portfolio slightly underperformed until 2012. Further, for the rolling window of 500 days we can note that the Clayton copula based portfolio, performed well during the initial crisis but started to lag behind when the turnaround came. This characteristic is particularly clear after the financial crises in 2007/08. If we compare the results between the two calibration windows we note that the total return was higher when we used a rolling window of 1250 observations compared to 500 observations.
Results

Rolling calibration window: 500 observations.

Figure 15: Backtest of the min CVaR portfolios for a rolling window of 500 days.

Rolling calibration window: 1250 observations.

Figure 16: Backtest of the min CVaR portfolios for a rolling window of 1250 days.
Expected return level: $r \geq 0.01\%$

In this part we present the results for the minimum CVaR portfolios, given that $r \geq 0.01\%$. See the red circle in Figure 13 for an example of the selected portfolio. From Figure 17 we see, like the previous graphs, that it was almost an equal performance until 2012 when they started to spread apart. Noteworthy is that, when we used a calibration window of 500 days, Markowitz portfolio this time outperformed the others from 2012.

Similar to the previous results, we see that the equally weighted portfolio again slightly underperformed before the final recovery starting from 2012. Figure 17 also shows that the Clayton copula based portfolio managed the initial crisis well but started to lag behind in the last upturn. In a comparison between the two length of calibration windows, we note once again that the total return, in general, was higher for the rolling window of 1250 days compared to 500 days.

Rolling calibration window: 500 observations.

![Figure 17: Backtest of portfolios with $r \geq 0.01$ for a rolling window of 500 days.](image-url)
Results

Rolling calibration window: 1250 observations.

![Portfolio performance graph](image)

*Figure 18: Backtest of portfolios with \( r \geq 0.01 \) for a rolling window of 1250 days.*

**Expected return level:** \( r \geq 0.02\%

In this part we present the results for the minimum CVaR portfolios, given that \( r \geq 0.02\% \). See the red circle in Figure 13 for an example of the selected portfolio. From these results, shown in Figure 19 and 20, we can see that there were once again, almost an equal performance for the portfolios until 2012.

From this optimization we see that Markowitz portfolio performed slightly better than the others, with a calibration window of 500 days. In the same way we also note that the Student-t portfolio performed better than the other portfolios when we used a calibration window of 1250 days. When comparing the two length of calibration windows, the rolling window of 1250 yielded, in general, a slightly higher total return compared to 500 days.
Results

Rolling calibration window: 500 observations.

Figure 19: Backtest of portfolios with $r \geq 0.02$ for a rolling window of 500 days.

Rolling calibration window: 1250 observations.

Figure 20: Backtest of portfolios with $r \geq 0.02$ for a rolling window of 1250 days.
Expected return level: $r \geq 0.03\%$

In this part we present the results, shown in Figure 21 and 22, for the minimum CVaR portfolios, given that $r \geq 0.03\%$. See the red circle in Figure 13 for an example of the selected portfolio.

From these graphs, it is worth noting that Markowitz portfolio performed well when we used the calibration window of 500 days whereas the Student-t portfolio seems to have done well for a long period when we used a calibration window of 1250 days. Like the previous results, the equally weighted portfolio underperformed during several periods, although it was not by much. Furthermore, the graphs indicates that the total return for the Clayton copula was significantly higher for the rolling window of 1250 days compared to 500 days.

Rolling calibration window: 500 observations.

![Graph](image)

*Figure 21: Backtest of portfolios with $r \geq 0.03$ for a rolling window of 500 days.*
Results

Rolling calibration window: 1250 observations.

Figure 22: Backtest of portfolios with $r \geq 0.03$ for a rolling window of 1250 days.

Performance metrics
To evaluate the performance during the entire period we need more than graphs. Thus, below follows six different types of performance metrics that are chosen to reflect the portfolios risk or return in different perspectives.

In Table 3 we see the total return for each portfolio by the end date, 22 February 2016. The total return is defined as the change in portfolio value during the backtesting period. This measure should not be taken to seriously since, as stated before, focusing to narrowly on one date will have major impact on conclusions. However, this metric still gives us some information about the performance.

From this table we note that the use a rolling window of 1250 days generally yielded a higher total return compared to 500 days, with some few exceptions. As we have seen in the graphs above, Markowitz portfolio performed well for the calibration window of 500 days while for 1250 days the results were tighter. Moreover, we note that the total return generally increased as the level of expected return increased.
Table 3: Total return for each portfolio, rolling calibration window and required expected of return.

<table>
<thead>
<tr>
<th>Window</th>
<th>Total return (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 days</td>
</tr>
<tr>
<td>Expected return</td>
<td>min CVaR</td>
</tr>
<tr>
<td>Normal</td>
<td>44.55</td>
</tr>
<tr>
<td>Gaussian</td>
<td>40.12</td>
</tr>
<tr>
<td>Student-t</td>
<td>35.22</td>
</tr>
<tr>
<td>Clayton</td>
<td>20.74</td>
</tr>
<tr>
<td>Equal</td>
<td>45.05</td>
</tr>
</tbody>
</table>

Table 4 below shows the average daily return for each portfolio during the backtesting period. This metric is more important than the total return since it takes into account the entire period. The result indicates that, in general, the average daily return increased as the expected return increased. According to this table, Markowitz portfolio performed better than the other portfolios when we used the rolling window of 500 days. Though, there was a wider range of the results for the rolling window of 1250 days. We also note that except for Markowitz portfolio, the rolling calibration window of 1250 days increased the average daily return.

Table 4: Average daily return for each portfolio, rolling calibration window and required expected of return.

<table>
<thead>
<tr>
<th>Window</th>
<th>Average daily return (10^{-2})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500 days</td>
</tr>
<tr>
<td>Expected return</td>
<td>min CVaR</td>
</tr>
<tr>
<td>Normal</td>
<td>0.0150</td>
</tr>
<tr>
<td>Gaussian</td>
<td>0.0145</td>
</tr>
<tr>
<td>Student-t</td>
<td>0.0135</td>
</tr>
<tr>
<td>Clayton</td>
<td>0.0113</td>
</tr>
<tr>
<td>Equal</td>
<td>0.0165</td>
</tr>
</tbody>
</table>
In Table 5 we see the daily volatility of each portfolio. This metric will give us an indication about the risk and even though this risk metric has been criticized, we still use it as a performance indicator due to its simplicity. From the previous table we could see that the rolling window of 1250 days yielded a higher average daily return. This can be explained by a higher level of risk since the results from Table 5 suggests that the volatility in general was higher for 1250 days compared to 500.

When comparing the portfolios results against each other we note that the Student-t copula based portfolio, in general had the lowest volatility. Furthermore, the equally weighted portfolio had in general higher risk when we used 500 days. Though, for 1250 days it seems, except for the min CVaR portfolio, that only the Student-t copula based portfolio beat it.

Table 5: The volatility defined by the standard deviation for each portfolio, rolling calibration window and minimum required level of return.

<table>
<thead>
<tr>
<th>Window</th>
<th>500 days</th>
<th>1250 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min CVaR</td>
<td>0.01</td>
</tr>
<tr>
<td>Normal</td>
<td>10.06</td>
<td>10.92</td>
</tr>
<tr>
<td>Gaussian</td>
<td>10.24</td>
<td>10.95</td>
</tr>
<tr>
<td>Student-t</td>
<td>10.22</td>
<td>10.62</td>
</tr>
<tr>
<td>Clayton</td>
<td>11.15</td>
<td>10.97</td>
</tr>
<tr>
<td>Equal</td>
<td>11.25</td>
<td></td>
</tr>
</tbody>
</table>

Table 6 shows the maximum loss for each the portfolio over all trading days. This metric is somehow coupled with the volatility since it measures the risk taken but in a different perspective. It is notable that, unlike the average volatility where the risk was higher for 1250 days, with metric the risk was lower for the same window since it in general led to a slightly lower maximum draw down.

However, like the previous results for volatility, the Student-t copula based portfolio exhibited a good performance for 1250 days. The equally weighted portfolio had overall a good performance on maximum draw down together with the Student-t and sometimes the Gaussian copula. Furthermore, we can note that Markowitz portfolio performed well for the rolling window of 500 days.
### Maximum draw down (%)

<table>
<thead>
<tr>
<th>Window</th>
<th>500 days</th>
<th>1250 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min CVaR</td>
<td>0.01 0.02 0.03</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>6.85</td>
<td>6.93 6.93 6.93</td>
</tr>
<tr>
<td>Student-t</td>
<td>5.76</td>
<td>7.49 7.49 7.49</td>
</tr>
<tr>
<td>Clayton</td>
<td>6.69</td>
<td>7.43 7.43 7.43</td>
</tr>
</tbody>
</table>

The two last and simplest, but yet so useful performance metrics are the number of months with highest and lowest return. The results are presented in Table 7 and 8 and from them we note that the equally weighted portfolio had the highest number of months with both best and worst performance. We also note that Markowitz portfolio and especially the Gaussian copula performed well for both length of calibration windows.

### Number of month with highest returns

<table>
<thead>
<tr>
<th>Window</th>
<th>500 days</th>
<th>1250 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min CVaR</td>
<td>0.01 0.02 0.03</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaussian</td>
<td>27</td>
<td>33 33 30</td>
</tr>
<tr>
<td>Student-t</td>
<td>21</td>
<td>14 14 17</td>
</tr>
<tr>
<td>Clayton</td>
<td>20</td>
<td>30 27 26</td>
</tr>
<tr>
<td>Equal</td>
<td>40</td>
<td>33 35 34</td>
</tr>
</tbody>
</table>

57
Table 8: Number of month with lowest return for each portfolio, rolling calibration window and a required expected of return.

<table>
<thead>
<tr>
<th>Window</th>
<th>Number of month with lowest returns</th>
<th>500 days</th>
<th></th>
<th></th>
<th></th>
<th>1250 days</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>min CVaR</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>min CVaR</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>Normal</td>
<td></td>
<td>21</td>
<td>19</td>
<td>19</td>
<td>18</td>
<td>24</td>
<td>19</td>
<td>28</td>
<td>29</td>
</tr>
<tr>
<td>Gaussian</td>
<td></td>
<td>16</td>
<td>18</td>
<td>20</td>
<td>19</td>
<td>32</td>
<td>23</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>Student-t</td>
<td></td>
<td>20</td>
<td>29</td>
<td>24</td>
<td>28</td>
<td>16</td>
<td>31</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Clayton</td>
<td></td>
<td>59</td>
<td>49</td>
<td>51</td>
<td>50</td>
<td>53</td>
<td>49</td>
<td>39</td>
<td>32</td>
</tr>
<tr>
<td>Equal</td>
<td></td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>50</td>
<td>40</td>
<td>43</td>
<td>47</td>
<td>48</td>
</tr>
</tbody>
</table>
5 Discussion

In the paper “Portfolio-Selection” from 1952, Harry Markowitz introduced modern portfolio theory (MPT). The main idea of MPT is to construct a portfolio that maximizes the expected return for a given level of risk or in the opposite way minimizes the risk given a level of expected return. Even though this theory has gained in popularity over the years and influenced management practice, it has been criticized for the assumptions that are required. A fundamental assumption of MPT is that the future asset returns are normally distributed. However, in the real-world we have observed that returns do not exhibit this behavior. This also implies that the use of linear correlation loses its meaning since linear correlation only works well with elliptical distributions such as the normal distribution. By introducing and modelling the dependence structure with copula functions we allow for selection of non elliptical distributions.

In this paper we have demonstrated one way of using copulas in a portfolio optimization framework using CVaR as a risk metric. The objective of this paper was twofold. First, we examined if the use of copula functions in portfolio optimization is more efficient compared to the traditional MPT. Second, we examined how the presence of memory when calibrating the copula functions affects the chosen portfolios. Further, we also explored which of the chosen copula functions that showed the best fit to the historical data.

We have used three different copulas: Gaussian, Student-t and Clayton, which were presented in section 2.3.3. For comparison we also chose Markowitz portfolio and an equally weighted portfolio. It is important to note that Markowitz portfolio is comparable with the copula based portfolios since the efficient frontier is the same for both CVaR and volatility when using normally distributed returns. The length of the calibration windows were set to 500 and 1250 days respectively which means that we in total backtested nine portfolios for each set of constraints. The performance of each portfolio was evaluated using a dynamic portfolio strategy where the weights were updated on a monthly basis from 2000-2016.

The results from this study can be seen in section 4 where we provide a comparison between the portfolios performance. When interpreting the result from a portfolio optimization comparison it is of great importance not focus too narrowly on the end date of the backtesting period. The selection of end date might have major impact on the total return and we have therefore evaluated the portfolios during the whole period by using different performance metrics. For example, the compounded return seems to be quite similar for all portfolios and calibration windows until 2012. This shows that if we only focus on December 2012 as the end date, this study would have led to other conclusions.

In general we can conclude that portfolio optimization can be used in a successful way compared to choosing an equally weighted portfolio, even though there are some exceptions. Our results suggests that choosing equal weights seems to be a better allocation method than using the one parameter Clayton copula. This can be explained by the poor fit of the Clayton copula function. In Appendix C, the Goodness of Fit (GoF) which
Discussion

is calculated by mean square error (MSE), indicates that the Clayton copula showed an inferior fit during nearly the entire period compared to the elliptical copulas. A possible reason might be that the Clayton copula only has one parameter describing the strength of dependence. However, this comes not as a surprise since we explained this in section 2.3.3. It is a well-known problem that constructions of one parameter Archimedean copulas in higher dimensions easily can fail. Compared to the other optimized copula based portfolios, the equally weighted portfolio had in general a higher risk for the shorter calibration window and a lower average daily return for the longer calibration window.

Moreover, the results indicates that the Student-t copula overall had low risk in terms of volatility and appears to be the most efficient copula for the rolling calibration window of 1250 days. However, for the minimum CVaR portfolio, other portfolios seems to be as good as the Student-t copula. One possible explanation for the good performance, when we used a longer calibration window, can be the that the Student-t copula manage to capture the non normal distribution of returns. This fact is supported by the estimated degrees of freedom, illustrated in appendix C. From Figure 29 we can see that degrees of freedom was significantly lower for 1250 days compared to 500 days which indicates that the returns differs from the normal distribution.

For the rolling window of 500 days, both Markowitz and the Gaussian copula appears to estimate the distribution of returns relatively good. Even though Markowitz portfolio sometimes outperformed the other from 2012 it is central not to draw too rash conclusions only based on this period. A possible explanation why the copula based portfolios performed relatively worse than Markowitz under this period can be that the copulas were based on EVT which requires a certain amount of data in order to estimate the tail distributions in an accurate way. However, using too many observations for calibration can imply that the data is old and does not describe the current market conditions. Consequently, there exists a trade-off between using a large sample of data for calibration and inaccurate data. The copula parameters, which are shown in Appendix C, shows that there was a distinct difference during calm and stressed periods depending on the length of calibration window used. The rolling window of 500 days indicates that the parameter shifted more quickly after the current market condition.

So does Copula beat linearity? From the results of this paper it is no easy answer to that question beyond ”it all depends”. For some sub-periods of the backtesting period and required expected returns the copula-based portfolios with EVT slightly outperformed Markowitz and for some they were beaten. Consequently, this paper do not present an explicit answer to the problem statement. Instead, the main conclusion we can draw from our result is that the concept of portfolio optimization as a whole is an adequate way to allocate assets with a risk-return trade off. Furthermore, our results did not exhibit an explicit answer of which length that was preferable. Thus, the choice of length of the calibration window is dependent on the objective and risk tolerance of the investor. 1250 observations implies in general a higher risk (with some exceptions for maximum draw down) which also can be compensated with a higher average daily return. The higher risk
Discussion

for longer windows is most likely based on that the optimized portfolios contains more information but a shorter window allows the optimized portfolios to faster adapt to the current market condition and hence lower the risk. Thus, the length of the calibration window chosen depends largely on the investors opinion of the future market conditions.

Further research within this area could involve a time-dependent model that handles both autocorrelation and heteroskedasticity such as the GARCH-model. Another possible extension is to use Archimedean copulas with more than one parameter or vine copulas for investigating high-dimensional Archimedean copulas in a more accurate way. Further, a more simple extension could be to conduct a similar model as presented in this paper but instead choose other parameters to stress. Finally, it would be of great importance to investigate if the algorithms can be implemented in a more efficient way since the optimization took a long time to run.
References

Articles


References


**Books**


References


Other

Appendix A: Illustration of data

The price and return series for the six indices: TSX Composite, CAC40, DAX, Nikkei225, FTSE100 and S&P500 during the entire sample period 950101 - 160222 are presented in Figure 23 to 25.

*Figure 23: Normalized price and return for the whole sample period: TSX Composite & CAC40.*
Figure 24: Normalized price and return for the whole sample period: DAX & Nikkei225.
Figure 25: Normalized price and return for the whole sample period: FTSE100 & S&P500.
Appendix B: Summary statistics of the data

The pairwise correlation, histograms and QQ-plot of each indices: TSX Composite, CAC40, DAX, Nikkei225, FTSE100 and S&P500 during the entire sample period 950101 - 160222 are presented in Figure 26 and 27.

Histogram

Figure 26: Pairwise correlation and histograms of historical returns.
QQ-plot
In Figure 27 below we present a QQ-plot over the chosen indices.

*Figure 27: QQ-plot of the different indices against normal distribution.*
Sample pairwise correlation
In Figure 28 below we illustrate how the dependence in terms of Pearson’s linear correlation between the indices changed over the backtesting period.

Figure 28: Pairwise correlation between the indices during the backtesting period
Appendix C: Illustration of estimated copula parameters

The copula parameters that were calibrated by the two rolling windows with 500 and 1250 observations of daily returns respectively are presented in Figure 29-31. In Figure 32-33 we see the mean square mean square error (MSE) for all copula functions.

![Estimation of degrees of freedom for Student-t copula](image)

*Figure 29: Estimated degrees of freedom for the Student-t copula function for the two length of rolling windows.*
Appendix C

Figure 30: The average correlation coefficient for the estimated correlation matrices for Gaussian and Student-t copula function for the two length of rolling windows.

Figure 31: Estimated alpha parameter for Clayton copula function for the two rolling windows.
Figure 32: Mean square error for the rolling window of 500 days.

Figure 33: Mean square error for the rolling window of 1250 days.
Appendix D: Portfolio weights

The portfolio weights during the entire backtesting period for each portfolio:

**Expected return level: min CVaR**

![Graphs showing portfolio weights for different copulas over different years](image)

Figure 34: Portfolio weights during the backtesting period for the min CVaR portfolios and the rolling window of 500 days.

![Graphs showing portfolio weights for different copulas over different years](image)

Figure 35: Portfolio weights during the backtesting period for the min CVaR portfolios and the rolling window of 1250 days.
Expected return level: \( r = 0.01 \)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig36}
\caption{Portfolio weights during the backtesting period for the \( r = 0.01 \) portfolios and the rolling window of 500 days.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig37}
\caption{Portfolio weights during the backtesting period for the \( r = 0.01 \) portfolios and the rolling window of 1250 days.}
\end{figure}
Appendix D

Expected return level: $r=0.02$

Figure 38: Portfolio weights during the backtesting period for the $r = 0.02$ portfolios and the rolling window of 500 days.

Figure 39: Portfolio weights during the backtesting period for the $r = 0.02$ portfolios and the rolling window of 1250 days.
Expected return level: $r=0.03$

Figure 40: Portfolio weights during the backtesting period for the $r = 0.03$ portfolios and the rolling window of 500 days.

Figure 41: Portfolio weights during the backtesting period for the $r = 0.03$ portfolios and the rolling window of 1250 days.