Tail dependence in financial data

Modelling dependence in dynamic factor models with copulas and extreme value theory

by

Thijs Nicolaas Schouten

to obtain the degree of Master of Science
in Applied Mathematics
at the Delft University of Technology,
Faculty of Electrical Engineering, Mathematics and Computer Science,
to be defended publicly on Monday August 28, 2017 at 3:30 PM.

Student number:	4249712	
Project duration:	September 18, 2016 – August 28, 2017	
Thesis committee:	Prof. dr. ir. C.W. Oosterlee,	TU Delft, supervisor
	Dr. D. Kurowicka,	TU Delft
	Dr. P.F.A. Tuijp,	Ortec Finance

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft Delft University of Technology

ORTEC FINANCE

Abstract

In this thesis we model extreme log-returns on economic variables and apply this to Ortec Finance's model. These extreme log-returns are relevant for risk management applications such as Value-at-Risk and other measures of tail risk. We use extreme value theory to simulate economic variables with the desired tail behaviour. We pay special attention to correlations between economic variables, since these tend to increase during financial crises. This suggest the possibility of tail dependence and we use copula theory to model behaviour similarly to what we observed historically.

We find that a single parameter, the tail index, can be used to model the tail behaviour of an economic variable. To model the tail dependence between economic variables we can also use a single parameter namely the tail dependence coefficient. We model the complete dependence structure with a semiparametric copula, such that the copula has the desired tail dependence coefficient, but also approximates the dependence outside the tails.

These techniques are applied in the context of vector autoregressive models, since these models are used to describe the statistical factors in Ortec Finance's Dynamic Scenario Generator, which generates future economic scenarios. We provide a first stylized indication on how these techniques could be applied in the context of Ortec Finance's model.

Preface

This thesis has been submitted as the final requirement for the Master of Science degree in Applied Mathematics at the Delft University of Technology. First of all, I would like to thank my supervisor Patrick Tuijp for his time, ideas, feedback, and guidance throughout the project. It was great to work with someone so capable and enthusiastic. Special thanks for letting me start the project on a part-time basis during my exchange with ETH Zurich, despite the disadvantages of distant contact and supervision. I would like to thank Kees Oosterlee as well for taking the time to thoroughly read my thesis before every meeting, his feedback and ideas that helped to progress. Moreover I would like to thank Dorota Kurowicka for being part of my thesis committee and taking the time to read through my thesis and give helpful feedback halfway the project.

I also want to thank some people who have contributed to the project in the broader sense. I am very grateful for having such friendly colleagues at Ortec Finance with whom I had interesting discussions on mathematics, finance, and life. Last but not least, I would like to thank my family for the wise lessons, care, and support during my entire studies.

Thijs Nicolaas Schouten
Delft, August 2017

List of notation and abbreviations

Below we give a list of the most important notation and abbreviations used in this thesis. This list applies throughout the thesis, unless specifically mentioned otherwise.

γ	tail index (of the upper tail if not mentioned which tail)
γ^L	lower tail index
γ^U	upper tail index
ρ	second order parameter
F	distribution function
$\overline{F} = 1 - F$	survival function
U	the inverse of $\frac{1}{F}$
$X_{i,n}$	i-th smallest value of $\{X_1, \cdots, X_n\}$
$\text{Cov}(X_1, X_2)$	covariance between X_1 and X_2
$\mathbb{E}[X]$	expectation of X
$\text{Var}(X)$	variance of X
VaR_p	the p Value-at-Risk
(V)AR(p)	(vector) autoregressive with p lags
(V)MA(q)	(vector) moving average with q lags
f_t	(vector of) factor(s) at time t
OLS	ordinary least square
WLS	weighted least square
C	copula
$\overline{C}$	survival copula
c	copula density
g	generator of Archimedean copula
R_i^j	rank dimension i of the j-th element
τ	Kendall's tau
ρ_S	Spearman's rho
α	significance level
TDC	tail dependence coefficient
λ_U	upper tail dependence coefficient (sometimes denoted by λ)
λ_L	lower tail dependence coefficient
$\lambda_{U,L}$	upper-lower tail dependence coefficient
$\lambda_{L,U}$	lower-upper tail dependence coefficient
MSE	mean squared error
AMSE	average mean squared error
$\forall$	for all
$\exists$	(there) exists

Contents

1 Introduction

In this master thesis we will analyse the way Ortec Finance models extreme log-returns on financial instruments and comes up with a different approach to handle this problem. Extreme returns are large losses or gains that occur with a low probability. For risk management applications we wish to estimate the probability of very large losses. To do this, the behaviour of returns of financial instruments has to be modelled and in many financial applications parametric models are used. The dependence between two financial variables could be modelled with a single correlation parameter. This parametric model, can be fitted using maximum likelihood or other estimators, and in this case it approximates the behaviour of financial instruments around its mean. It does, however, not describe the behaviour of the outliers. Therefore, this method could underestimate the probability of extreme events corresponding to these outliers. The literature shows us that many log-returns of financial instruments show so-called heavy-tailed behaviour, while models sometimes use light-tailed distributions (e.g. the normal distribution). These models can severely underestimate the probability of extremely negative returns, for example the probability of a financial crisis or a single bankruptcy.

This is aggravated by the fact that correlations between economic variables increase during financial crises, indicating that a single constant correlation parameter might not be enough to model the dependence observed in the market. It is therefore important to model the behaviour of a set of variables correctly, especially the lower tail, since this part of the distribution explains the behaviour in critical financial times. To describe this lower tail we will explain the concept tail dependence and introduce a technique to estimate the tail dependence coefficients, that are single parameters used to describe the dependence in the tails of a distribution.

1.1. Ortec Finance model

Ortec Finance uses a dynamic factors model to describe the behaviour of many financial variables. These financial variables can be described by a linear combination of factors, lagged factors and an individual error term. The dynamics of these factors is captured by a vector autoregressive model with one lag (VAR(1) model) and normally distributed errors. Afterwards the factors are adjusted to account for the skewness and the tails that are fatter than normal distribution tails, but extreme value theory and copulas are not used to capture the tail dependence.

At Ortec Finance, a distinction is made between core variables and regression variables. The regression variables depend on the core variables and for the core variables Ortec Finance uses an advanced model to describe their behaviour. In this model the core variables are first decomposed in a trend and various frequency components and the core variable is then simply the sum of the different components. These components are obtained via a frequency decomposition, which will not be elaborated on here. Each frequency component of an individual economic variable depends on some statistical factors. The number of factors can be different for every frequency component. The dynamics of these factors is captured by a vector autoregressive model with one lag. The errors are described by normal distributions and the components of the errors can be correlated. These factors are adjusted to account for the skewness and tails that are fatter than tails from a normal distribution. We will show that extreme value theory and copulas are valuable tools to model individual tails and tail dependence. Besides, we will give an indication on how Ortec Finance could implement this in their current modelling.

1.2. Goals of the project

The main goal of the project is to provide an indication of whether extreme value theory and copula theory can be used within the VAR framework, which is used to model the statistical factors.

- We wish to be able to accurately estimate the historical tail behaviour of an economic variable, in particular a statistical factor.
- We wish to be able to accurately estimate the historical dependence structure, including tail dependence, between sets of economic variables, especially statistical factors. We will use copulas for this.
- We wish accurately draw samples from this copula to generate scenarios that match historically observed behaviour.
- We wish to generate statistical factors with similar tail index and dependence structure as observed historically.

1.3. Organisation of thesis report

In Chapter 2 we apply extreme value theory to simulate univariate factors with tail behaviour similar to what we have observed historically. In this chapter we model the tails of distributions with a single parameter, the tail index. We compare different estimators for this parameter and choose a combination that will be used on the factors. In many tail index estimators we need to select a threshold. We introduce the intuitive stability method for this and show that this method works better in mean squared error (MSE) for some of the tail index estimators than other methods from the literature.

Since the factors themselves are modelled by a VAR model, we analyse the link between the tails of the factors and the error in the VAR model in Chapter 3. We show that, under some conditions, the tail index of factors equals the largest tail index of the errors if the errors are independent and identically distributed. We should therefore be careful when modelling the factors by a VAR model with independent and identically distributed errors, since we have no indication that all factors have the same tail index.

In Chapter 4 we pay special attention to correlations between economic variables. These correlations tend to increase during financial crises and my therefore not be modelled appropriately by a single constant parameter. This increasing dependence in stressed scenarios suggests the possibility of tail dependence and we use copula theory to capture the multivariate behaviour in such a way that we can include tail dependence. The tail dependence coefficient (TDC) is a measure for this tail dependence and we compare different methods to use this to estimate this parameter. We will select one method, which performs best on the test samples and use this for estimating the TDC between factors. The TDC estimator that we use has an upward bias and may give positive values when there is in fact no tail dependence present in the sample. Therefore we will also test for tail dependence using hypothesis testing. We come up with a copula estimation procedure which includes a similar dependence structure and TDCs as observed historically for the 2-dimensional setting. Using this copula we can generate scenarios that show the same dependence structure as historically observed. With a small introduction to vine copulas we give a promising direction for the extension to more dimensions.

These techniques for describing the dependence are applied in the context of VAR models in Chapter 5. Using a copula to describe the dependence in the errors gives a copula autoregressive (COPAR) model. The aim is to provide a first stylized indication of whether or not these techniques could be promising for the Ortec Finance model.

We will complete this thesis with a discussion of the results and make recommendations for further research on related topics in Chapter 6.

The most important results, proofs and theory can be found in the chapters. In the appendix we will give additional results, proofs and theory.

2

Extreme value theory

2.1. Introduction

For risk management purposes we sometimes wish to estimate the probability of the realisation of certain losses on a portfolio. To describe the log-returns of the portfolio we can use parametric models and we can calculate the probability of certain losses under the assumption that the model holds true. The left tail of distribution functions corresponds to these large losses and we are therefore interested in the tail behaviour of the log-returns. Extreme value theory is a valuable tool that we will use to describe the tails of distribution functions. We will apply it to estimate the probabilities of the occurrences of extreme events. Extreme value theory distinguishes three cases for the distribution function:

- The density has no tail, meaning that the density has a finite endpoint (e.g. a uniform distribution).
- The density has a light tail, meaning that the density decays exponentially or faster to zero, such that all moments exist (e.g. a normal distribution).
- The density has a heavy tail, meaning that the density decays polynomially to zero, meaning that only a finite number of moments exist (e.g. a Cauchy distribution).

Malevergne and Sornette (2006) explain that the log-return distributions of equities do not behave according to a normal distribution. They state that the returns should be modelled with heavy-tailed distributions.

In this chapter, we will first show the relevance of extreme value theory in financial settings. Next, we will define a measure for the fatness of the tail of a distribution, namely the tail index γ, which can be used to estimate the probability of extreme events. In the final part we will introduce estimators for this γ and analyse the performance of each estimator, based on samples from simulated data. Furthermore, we will use these estimators on financial data.

2.2. The need for extreme value theory

Without the use of extreme value theory the probability of extreme events is likely to be underestimated. As a result the risk on a portfolio could be underestimated. We will illustrate this in this section by an example, but first we will introduce a risk measure. A widely used measure for risk of returns on an investment portfolio is the Value at Risk (VaR), which is given in Definition 2.1.

Definition 2.1 (Value at Risk)**.** *For a portfolio with return X, the VaR_p is the value for which the probability of dropping below this value is p. Mathematically we write:*

$$\mathbb{P}(X \leq \mathrm{VaR}_p) = p.$$

In other words: there is a probability p that the return is below VaR_p.

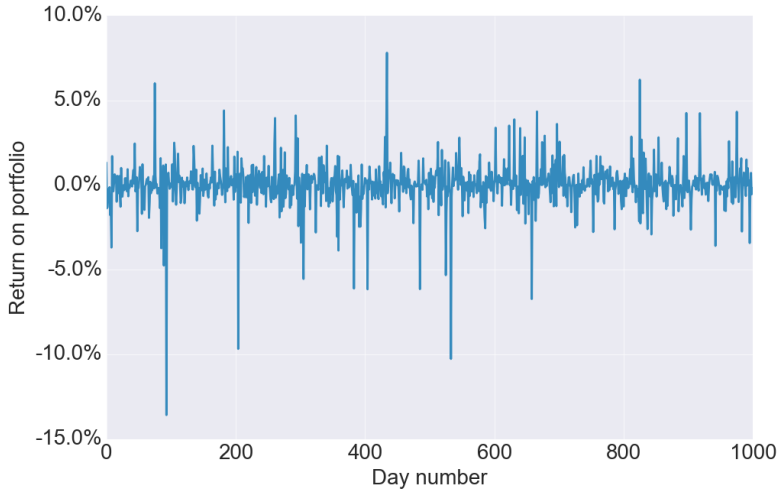
For a given sample size n we have that the lower p the more difficult to estimate VaR_p, since fewer events below this level have been observed for a lower p. Especially the $p < \frac{1}{n}$ case is difficult, since the corresponding events have never occurred in the historical data. The data in the centre of the distribution does not

say anything about how the distribution behaves in the tail and therefore how the extremes behave. These extremes are important, because these correspond to large losses on a portfolio. The following example illustrates how large the mistakes can be by using a naive approach without using extreme value theory.

Example 2.1. Consider a setting in which we wish to estimate the probability that a risky portfolio loses 20% of its value in one day (i.e. the return $X < -20\%$), based on historical data. Suppose we have data available for 1000 daily returns $X_1, \cdots, X_{1000}$, which are plotted in Figure 2.1. Since we simulated this data from a static distribution, we know the underlying distribution, which has no end points and two heavy tails. As we can see, the negative returns never exceed 20% and therefore counting the number of occurrences and dividing by the sample size will estimate the probability to be zero. We need to somehow extrapolate the distribution function to be able to say something useful about this probability.

Figure 2.1: **Log-returns on a portfolio**

This figure shows the simulated log-returns for a certain portfolio over 1000 days. The data is simulated from a double Pareto density.

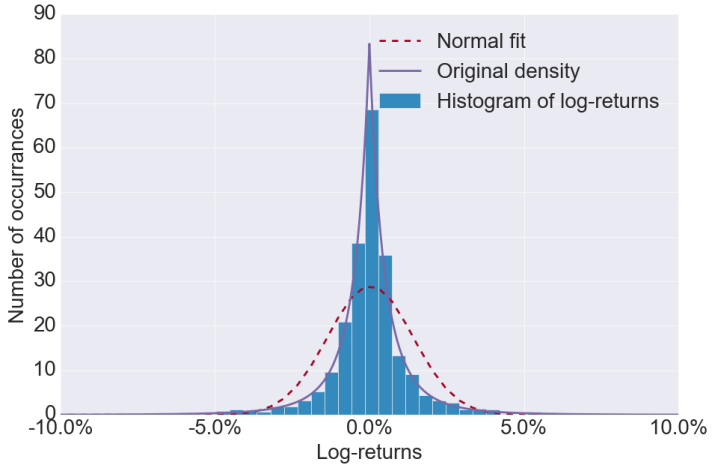


We could also fit a certain model, for example a normal distribution, to the data and estimate the parameters in the model. When we fit a normal density, we can calculate the variance and mean of the data and assume that the return of the portfolio behaves according to a normal distribution with this estimated mean and variance. Based on this normal distribution we can simply calculate the probability that $X < -20\%$. The downside of this approach is that we do not know what type of distribution X follows and we assume it to be of some parametric form. This could result in large errors when this assumption is not valid. In the case that we consider this is especially true in the tails of the distributions.

Suppose we take this approach and fit such a normal distribution to the daily returns $X_1, \cdots, X_{1000}$. This is shown in Figure 2.2, where we see that the normal distribution does not have the proper shape. Using this normal distribution we can now compute any probability $\mathbb{P}(X < -d)$ or $\mathbb{P}(X > d)$, so also the probability of losing 20%.

Figure 2.2: **Histogram and normal fit on simulated log-returns**

This figure gives a histogram of the simulated log-returns, a normal fit and the original density from which we drew the log-returns.



Using the normal estimate we could also write down the 4σ (99.994%) confidence interval, which is given by $[-0.0624, 0.0628]$. Under the normal assumption, these so-called 4σ-events (events outside this interval) occur once in 15787 data points (i.e. for 250 trading days per year this event would occur once every 63 years), but in the example it happens 9 out of 1000 times (i.e. just more than twice a year). The probability of such an event is then underestimated by a factor greater than 100, which becomes even worse when we look further in the tails, 5σ, 6σ, etc. events. It is clear that the assumption of the underlying process behaving according to a normal distribution is not correct, but we do not know which other distribution we should try. We could of course try other models, but these might also not describe the data in the tails as accurately as we wish. We therefore wish to avoid using a parametric approach that describes the whole distribution of the log-returns and this is where extreme value theory comes into play.

In all cases in which we want to estimate the probability of dropping below a certain value or conversely cases in which we want to estimate this value for a given probability we need some description of the tail. To describe this tail of distributions, often described by the tail index γ, we first introduce some notation. All the definitions are based on the right tail of the distribution. If we wish to know what happens in the left tail, which in financial applications correspond to large losses, we can simply consider $-X$ with the corresponding distribution and apply the definitions to $-X$. The definitions and propositions of this section, including further details, can be found in Haan and Ferreira (2006).

Definition 2.2 (Domain of attraction)**.** *F is in the domain of attraction of extreme value distribution G_γ if and only if there exist sequences $a_n \in \mathbb{R}_{>0}$ and $b_n \in \mathbb{R}$ such that $\forall x$ with $1+\gamma x > 0$:*

$$\lim_{n\to\infty} F^n(a_n x + b_n) = G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma x)^{-\frac{1}{\gamma}}\right), & \text{if } \gamma \neq 0 \\ \exp(-\exp(-x)), & \text{if } \gamma = 0 \end{cases}, \tag{2.1}$$

where $\gamma \in \mathbb{R}$. We write $F \in \mathscr{D}(G_\gamma)$.

When for the distribution function of X we have $F \in \mathscr{D}(G_\gamma)$, we say that X has tail index γ. If $\gamma > 0$ there are several ways to determine γ, which leads to the following proposition, see Haan and Ferreira (2006).

Proposition 2.1. *For X a random variable from distribution function F and $\gamma > 0$ the following statements are equivalent:*
(1) $F \in \mathscr{D}(G_\gamma)$
(2) X has tail index γ.
(3) $x^+ = \sup\{x : F(x) < 1\} = \infty$ and for $x > 0$:

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\frac{1}{\gamma}}. \tag{2.2}$$

(4) $x^+ = \sup\{x : F(x) < 1\} = \infty$ *and for* $x > 0$*:*

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \tag{2.3}$$

where U *is the inverse of* $\frac{1}{1-F}$*.*
(5) For x*:*

$$\mathbb{P}(X > x) \to Ax^{-\frac{1}{\gamma}} + \mathcal{O}\left(x^{-\beta}\right), \tag{2.4}$$

for some $\beta > \frac{1}{\gamma}$*.*

We will not give the proof here, but we will use the result. If $\gamma > 0$, referred to as the Fréchet domain, then the underlying distribution has a heavy right tail and the right endpoint x^+ is infinite. This also means that the moments of order greater than $\frac{1}{\gamma}$ do not exist (e.g. student's t-distribution, Cauchy distribution).

Other examples of distributions with a positive tail index are the Pareto, Fréchet and Burr distribution, given in Tadikamalla (1980). Later we will use samples from these distributions as test data to determine how well each estimator performs. The densities and distributions function are given in Table 2.1.

Table 2.1: **The density and distribution functions of the Pareto, Fréchet and Burr distribution.**

This table shows the density and distribution functions of the Pareto, Fréchet and Burr distribution with scale parameter 1 and domain $(0,\infty)$.

Type	Density	Distribution
Pareto	$f(x) = \frac{1}{\gamma}(x+1)^{-\gamma^{-1}-1}$	$F(x) = 1-(x+1)^{-\gamma^{-1}}$
Fréchet	$f(x) = \frac{1}{\gamma}x^{-\gamma^{-1}}e^{-x^{-\gamma^{-1}}}$	$F(x) = e^{-x^{-\gamma^{-1}}}$
Burr	$f(x) = \frac{1}{\gamma}x^{-\gamma^{-1}-1}\left(1+x^{-\gamma^{-1}}\right)^{-2}$	$F(x) = \left(1+x^{-\gamma^{-1}}\right)^{-1}$

All these distributions have no lower tail, since all densities have support $(0,\infty)$. The upper tail index is given by parameter $\gamma > 0$.

When a distribution of a random variable is not heavy-tailed, we can either use Proposition 2.2 or 2.3 to compute the tail index, see Haan and Ferreira (2006).

Proposition 2.2. *A random variable* X *with distribution function* F *has a light (upper) tail, meaning* $\gamma = 0$*, if and only if there exists a positive function* f*, such that:*

$$\lim_{t\uparrow x^*} \frac{1-F(t+xf(t))}{1-F(t)} = e^{-x}, \tag{2.5}$$

where x^* *can be either finite or infinite. We refer to this as the Gumbel domain. All moments for these distributions exist (e.g. normal distribution, gamma distribution).*

Proposition 2.3. *A random variable* X *with distribution function* F *has no upper tail, meaning* $\gamma < 0$*, if and only if:*

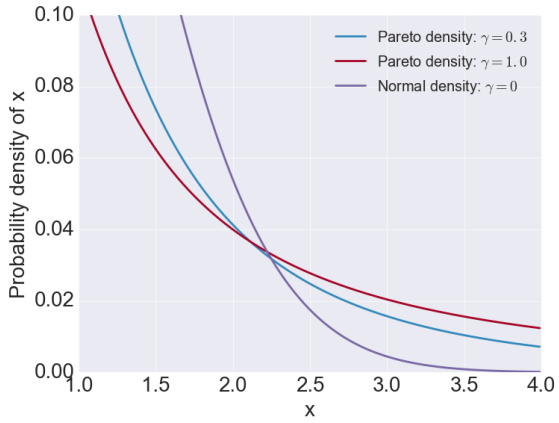
$$\lim_{t\downarrow 0} \frac{1-F(x^*-tx)}{1-F(x^*-t)} = x^{-\frac{1}{\gamma}}, \tag{2.6}$$

where $x^* < \infty$ *(i.e. F has a finite endpoint). We refer to this as the reverse-Weibull domain (e.g. uniform distribution).*

The tail index gives information about the speed of convergence of the density function towards zero. Figure 2.3 shows the tails of three density functions with different corresponding tail indices.

Figure 2.3: **Tails of three different density functions.**

This figure show the right tails of a normal density and two Pareto distributions.



Of the three densities in Figure 2.3, the normal density converges fastest to zero, since it has the lowest tail index. The other two have a positive tail index and therefore decay polynomially to zero instead of exponentially, for these two distributions not all moments exist.

The tail index γ is a first order parameter which shows the fatness of the tail. It does, however, not say anything about the speed of convergence towards this tail. Hence, we introduce the second order parameter in Definition 2.3, see Haan and Resnick (1996).

Definition 2.3. *For a heavy-tailed distribution with tail index γ the second order parameter is the $\rho \leq 0$ for which the following holds:*

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = \begin{cases} x^{\gamma}\frac{x^{\rho}-1}{\rho} & \text{if } \rho \neq 0 \\ x^{\gamma}\log x & \text{if } \rho = 0 \end{cases}, \quad \text{for some } A(t) \to 0 \ (t \to \infty). \tag{2.7}$$

When we wish to estimate probabilities of extreme events, estimating the tail index γ is important. To improve this estimate, some estimators also include an estimate of this ρ.

Now suppose we have sample $\{X_1, ...X_n\}$ where all X_i's independently follow the same distribution as X and we wish to estimate x_p, where $\mathbb{P}(X > x_p) = p$ for some p close to zero. Note that this is equivalent to estimating VaR_p. Haan and Ferreira (2006) show us how we can do this by applying extreme value theory in which case we do not have to assume any parametric form of the underlying distribution. First we will combine equation (2.3) with the fact that $x_p = U\left(\frac{1}{p}\right)$ to obtain the following estimate.

$$x_p = U\left(\frac{1}{p}\right) \approx U\left(\frac{n}{k}\right)\left(\frac{k}{np}\right)^{\gamma}, \tag{2.8}$$

which is an accurate approximation for $\frac{n}{k}, k \gg 1$ and here we have to choose the threshold k. We continue by introducing the ordered sample of $\{X_1, ..., X_n\}$, which is given in Definition 2.4.

Definition 2.4 (Ordered sample)**.** *Let $\{X_1, ..., X_n\}$ be a sample. Then the ordered sample is written as $\{X_{1,n}, X_{2,n}, ..., X_{n,n}\}$ where $X_{1,n} \leq X_{2,n} \leq ... \leq X_{n,n}$ and $\{X_1, ..., X_n\} = \{X_{1,n}, X_{2,n}, ..., X_{n,n}\}$.*

We can estimate $U\left(\frac{n}{k}\right)$ by $X_{n-k,n}$ and obtain the following estimator for x_p:

$$x_p \approx \hat{x}_p = X_{n-k,n}\left(\frac{k}{np}\right)^{\gamma}. \tag{2.9}$$

Here the tail index γ is still unknown and we therefore proceed with estimation methods for this γ in the next section.

2.3. Tail index estimation

In this section we will first introduce the relatively simple Hill and Moment estimators. It is know that these estimators have a bias (see Haan and Ferreira (2006)). This bias increases with the threshold and we will elaborate on this in the next section. Other more advanced estimators, that correct for the bias of these estimators are introduced next: the Regression and adjusted Hill estimator. Subsequently, we will show ways to determine the optimal threshold for all the estimators and show the performance of the estimators on simulated data sets with different tail indices, the test data. We will finally consider all linear combinations with equal weight of the four estimators and select the one which performs best on the test data. This best linear combination will be used throughout the rest of this thesis.

The estimation of the tail index γ is not an easy task. This index describes the first order behaviour of the tail of the distribution, but in a finite sample it can be the case that only a few observations are representative for this tail. For describing the behaviour of the tail we only look at the largest values (or smallest when dealing with the lower tail). Taking many observations into account for estimating this tail behaviour usually goes hand in hand with a large bias, since more observations will be outside the tail when we look at more observations. On the other hand, taking too few observations into account gives an estimate with a larger standard error. Choosing the right number of data points used for the estimate and methods to reduce the bias are important in making an accurate estimation for γ.

2.3.1. The Hill and Moment estimator

The Hill estimator is the most well-known estimator for the tail index and defined by Definition 2.5.

Definition 2.5 (Hill estimator)**.** *Suppose we have a sample* $\{X_1, \cdots, X_n\}$ *of which* $\{X_{1,n}, \cdots, X_{n,n}\}$ *is the ordered sample, then the Hill estimator is given by:*

$$\hat{\gamma}^H = \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right) = \frac{1}{k}\sum_{i=1}^{k}\log\frac{X_{n-i+1,n}}{X_{n-k,n}}, \tag{2.10}$$

where we see that the estimator is a function of the threshold k*.*

In general we wish to have k large such that we have a sufficient number of observations. Also we wish $\frac{k}{n}$ to be small, such that we have many tail observations relative to observations from the centre of the distribution, since those in the centre are not representative for the tail behaviour. The behaviour of the Hill estimator is discussed by Haan and Ferreira (2006) and we summarize this by the following two theorems. The proofs of both theorems are also given in Haan and Ferreira (2006) and will not be elaborated on here.

Theorem 2.4. *For any random variable* X *with tail index* $\gamma > 0$*,* k *fixed and* $n \to \infty$ *we have that:*

$$\sqrt{k}(\hat{\gamma}^H - \gamma) \xrightarrow{d} \mathcal{N}(0, \gamma^2), \tag{2.11}$$

so the Hill estimator is asymptotically consistent.

In the estimations we deal with a finite sample and therefore we need to choose an appropriate threshold. Choosing k large gives us many data points resulting in a small variance, but it also means that we have more observations outside the tail leading to a larger bias. Theorem 2.5 shows the trade-off we have to make here.

Theorem 2.5. *For any random variable* X *with tail index* $\gamma > 0$ *and* $\hat{\gamma}^H$ *based on sample* $\{X_1, ..., X_n\}$*, where* k *is chosen as threshold, the following holds:*

$$\hat{\gamma}^H - \gamma \xrightarrow{d} \mathcal{N}\left(\frac{\mu(k)}{k}, \frac{\gamma^2}{k}\right), \tag{2.12}$$

where the bias $\frac{\mu(k)}{k}$ *for which the* μ *function is unknown, increases in* k *and the standard error* $\frac{\gamma}{\sqrt{k}}$ *is decreasing in* k*.*

Since $X_{n-i+1,n} \geq X_{n-k,n} \forall i \leq k$ all the logs are positive and the Hill estimator cannot be negative. This is an issue when we try to estimate the tail index for data generated from a distribution with $\gamma < 0$. It can be shown that $\hat{\gamma}^H$ estimates $\gamma_+ = \max\{0, \gamma\}$. If we find an estimator for $\gamma_- = \min\{0, \gamma\}$, then we can use the fact that $\gamma = \gamma_+ + \gamma_-$ to obtain a general estimate for $\gamma \in \mathbb{R}$. Now define the log-moments of the sample as:

$$M_n^{(j)} = \frac{1}{k}\sum_{i=1}^{k}\left(\log\frac{X_{n-i+1,n}}{X_{n-k,n}}\right)^j. \tag{2.13}$$

Then Haan and Ferreira (2006) discuss that $1-\frac{1}{2}\left(1-\frac{(\hat{\gamma}^H)^2}{M_n^{(2)}}\right)^{-1}$ is an accurate estimate for γ_- and this leads to the Moment estimator.

Definition 2.6 (Moment estimator)**.** *Suppose we have a sample* $\{X_1,\cdots,X_n\}$ *of which* $\{X_{1,n},\cdots,X_{n,n}\}$ *is the ordered sample, then the Moment estimator is given by:*

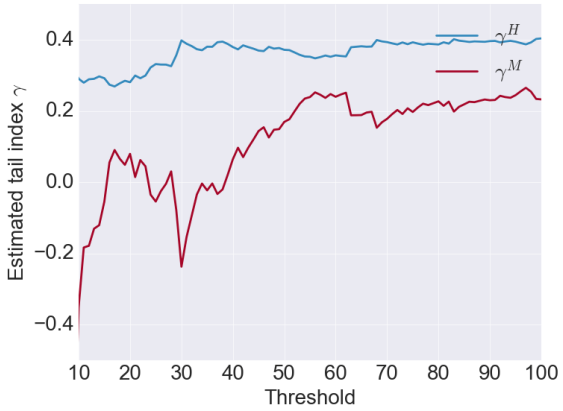
$$\hat{\gamma}^M = \hat{\gamma}^H + 1 - \frac{1}{2}\left(1-\frac{\left(\hat{\gamma}^H\right)^2}{M_n^{(2)}}\right)^{-1}, \tag{2.14}$$

where $\hat{\gamma}^H$ *and* $M_n^{(2)}$ *can be computed using equations (2.10) and (2.11). This is again a function of a threshold* k*.*

This estimator corrects the Hill estimator and has a smaller bias when we are dealing with small γ, close to and below 0. As an example we show the estimates of the tail index of the log-returns on the S&P 500, using the Hill and Moment estimators. We should realise here that both estimators work best for independent samples from a static distribution. If there is autocorrelation the estimator could perform worse. Furthermore the data might not follow a static distribution.

Figure 2.4: **Hill and Moment estimator on S&P 500 data**

This figure shows the Moment and Hill estimator for the daily returns of the S&P 500 in the period between 2006 and 2016.



In Figure 2.4 we see that the estimation for γ depends strongly on the choice of the threshold k. Both estimators show large fluctuations for thresholds $k<10$. As k increases these fluctuations become smaller, because the variance of the estimators decrease with k. We should however realise that the bias increases in k, which cannot be directly observed in this plot. For the Moment estimator the fluctuations are larger, but the bias smaller. The optimal threshold is therefore usually larger for this Moment estimator.

We can choose the threshold in several ways, that will be discussed in more detail in section 2.4. One way to do this is to check where the tail index is stable for the first time with respect to the threshold. For the Hill estimator we would choose a threshold of $k\approx 55$ corresponding to $\gamma\approx 0.34$ and for the Moment estimator $k\approx 88$ corresponding to $\gamma\approx 0.23$. The estimators have that $\gamma^M<\frac{1}{4}<\frac{1}{3}<\gamma^H$ and if the Hill estimator with this threshold gives the true value of γ we should conclude that the 3-rd and 4-th moment of the log-returns on the S&P 500 do not exist. If we suppose that the Moment estimator gives the true value of γ we should conclude that these moments do exist, which is a completely different result. It is therefore important to come up with a threshold selection method that results in accurate estimates and to choose the best estimator for the problem.

There exist many estimators that try to improve on the Hill estimator. We can for example take even higher order log-moments ($M_n^{(j)}$ for $j>2$) into account to try and improve on the Hill estimator. Considering these estimators can possibly improve our estimates. We can also take the second order parameter ρ into account to improve the estimate, recall that ρ is given in Definition 2.3. This is done in the Regression estimator and adjusted Hill estimator, given in sections 2.3.2 and 2.3.3.

2.3.2. The Regression estimator

To be able to come up with the Regression estimator we first rewrite the Hill estimator in equation (2.15).

$$\hat{\gamma}^H = \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n} = \frac{1}{k}\sum_{i=1}^{k} i \log \frac{X_{n-i+1,n}}{X_{n-i,n}} = \frac{1}{k}\sum_{i=1}^{k} y_i. \tag{2.15}$$

Beirlant et al. (1999) show that y_i approximately behaves according to an exponential distribution:

$$y_i = i \log \frac{X_{n-i+1,n}}{X_{n-i,n}} \overset{\mathrm{d}}{\approx} \left(\gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho}\right) g_i, \tag{2.16}$$

where $\rho < 0$ the second order parameter, $b_{n,k}$ unknown but does not depend on i and g_i are independent and identically distributed standard exponential random variables. We can substitute $e_i := g_i - 1$, where e_i is independent and identically distributed with zero-mean and obtain the following:

$$y_i = \gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho} + \left(\gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho}\right) e_i. \tag{2.17}$$

Beirlant et al. (1999) use an ordinary least square (OLS) regression to minimize the mean squared error and obtain estimates for γ and $b_{n,k}$ of which we are especially interested in the tail index γ. In equation (2.18) we write y in a different form.

$$y = Ax + \epsilon, \tag{2.18}$$

where $y = (y_1, ..., y_k)^T$, $x = (\gamma \;\; b_{n,k})^T$ and

$$A = \begin{pmatrix} 1 & \left(\frac{1}{k+1}\right)^{-\rho} \\ 1 & \left(\frac{2}{k+1}\right)^{-\rho} \\ \vdots & \vdots \\ 1 & \left(\frac{k}{k+1}\right)^{-\rho} \end{pmatrix}.$$

Here, the error vector ϵ is component-wise independent, but heteroscedastic (i.e. the error does not have the same variance for each component). The variance of the errors, given by $\left(\left(\gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho}\right)\right)^2$, is namely a function of i. In case of independent and identically distributed random variables we could simply use an ordinary least squares (OLS) regression which is given by:

$$\left(\hat{\gamma}^{(1)} \;\; \hat{b}^{(1)}_{n,k}\right)^T = \hat{x}^{(1)} = (A^T A)^{-1} A^T y. \tag{2.19}$$

Since the errors do not have the same variance for each component we can use some weighting such that the errors with larger variance contribute less to the estimate and improve the estimate. The first regression step gives an initial guess of γ and $b_{n,k}$. We therefore have an indication of the variances of the errors and we can rescale using this variance such that all error components have approximately the same variance, given that the initial estimate is close to the true values of γ and $b_{n,k}$. Using the rescaled values we can again do the regression. When we repeat this procedure we end up with the weighted least squares (WLS) method and more details can be found in Fahrmeir et al. (2007). This results in the following:

$$\frac{y_i}{c_i^{(1)}} \approx \frac{\gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho}}{c_i^{(1)}} + \frac{\left(\gamma + b_{n,k}\left(\frac{i}{k+1}\right)^{-\rho}\right)}{c_i^{(1)}} e_i, \tag{2.20}$$

where $c_i^{(1)} = \gamma^{(1)} + b_{n,k}^{(1)}\left(\frac{i}{k+1}\right)^{-\rho}$. This gives us:

$$\left(\hat{\gamma}^{(2)} \;\; \hat{b}^{(2)}_{n,k}\right)^T = c_i^{(1)} \hat{x}^{(2)} = c_i^{(1)} \left(A_1^T A_1\right)^{-1} A_1^T y^{(1)}, \tag{2.21}$$

where $y^{(1)} = \frac{y}{c_i^{(1)}}$, $A_1 = \frac{1}{c_i^{(1)}} A$. This procedure can be repeated for $c_i^{(2)}, c_i^{(3)}, ..., c_i^{(n)}$ with stop condition $|\gamma^{(n)} - \gamma^{(n-1)}| < \delta$ for some small δ. In the cases that we have considered, with $\delta = 0.01$, three or four iterations suffice, since the γ estimates do not change much. For a given k and y this can be summarized in Algorithm 2.1.

Algorithm 2.1 (Regression estimator)**.**

$y^{(0)} = y$
$A^{(0)} = A$
$b = 0, \gamma_0 = 0,\ \gamma_1 = 1$

while $|\gamma_1 - \gamma_0| < \delta$: **do**
  $c_i = \gamma_1 + b\left(\frac{i}{k+1}\right)^{-\rho}\ \forall 1 \le i \le k$
  $y_i = y_i^{(0)} / c_i\ \forall 1 \le i \le k$
  $A_i = A^{(0)} / c_i\ \forall 1 \le i \le k$ (update the rows of A)
  $\gamma_0 = \gamma_1$
  $[\gamma_1, b] = \left(A^T A\right)^{-1} A^T y_i$
end while
return γ_1

Here $y^{(0)}$ and $A^{(0)}$, first two lines of the algorithm, are the same as in equation (2.18). The second order parameter ρ is usually not known and there are several ways to deal with this. The estimate of γ turns out to be not so dependent on the choice of ρ and Vandewalle et al. (2004) suggested to take $\rho = -1$ or $\rho = -\frac{1}{2}$. One could also estimate ρ first and then use the estimate of ρ in the regression. This estimation procedure for ρ will be discussed in section 2.3.4.

2.3.3. Adjusted Hill estimator

Just like the Regression estimator, the Adjusted Hill estimator uses the second order parameter ρ to correct for the bias, which is present in the Hill estimator. It does, however, assume a parametric form of the underlying data, which holds true for the most used heavy-tailed distributions, for example for the Student's t and Fréchet distributions this is satisfied. However, for an arbitrary financial instrument this does not need to be the case, so we should be careful when using this estimator on arbitrary financial data. The assumption is that for some $\beta \neq 0$ and $x \to \infty$:

$$\mathbb{P}(X > x) \to \left(\frac{x}{C}\right)^{-\frac{1}{\gamma}} \left(1 + \frac{\beta}{\rho}\left(\frac{x}{C}\right)^{\frac{\rho}{\gamma}} + \mathcal{O}\left(x^{\frac{\rho}{\gamma}}\right)\right), \tag{2.22}$$

where ρ is defined above and $C > 0$. Based on equation (2.22) Gomes and Martins (2002) introduced the Adjusted Hill estimator. Since the assumption is not valid for $\gamma \le 0$ we should be careful when using this estimator. This estimator is given in Definition 2.7 and for statistical behaviour we refer to Gomes and Martins (2002).

Definition 2.7 (Adjusted Hill estimator)**.** *Suppose we have a sample* $\{X_1, \cdots, X_n\}$ *of which* $\{X_{1,n}, \cdots, X_{n,n}\}$ *is the ordered sample, then the Adjusted Hill estimator is given by:*

$$\hat{\gamma}^A = \hat{\gamma}^H \left(1 - \frac{\hat{\beta}}{1-\hat{\rho}}\left(\frac{n}{k}\right)^{\hat{\rho}}\right) \tag{2.23}$$

.

Here we need to find estimators $\hat{\rho}$ and $\hat{\beta}$ for ρ and β. Gomes and Martins (2002) also come up with methods to estimate these parameters. We will start with the estimator of β, which is given by equation (2.24).

$$\hat{\beta}(k) := \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{\left(\frac{1}{k}\sum_{i=1}^k \left(\frac{i}{k}\right)^{-\hat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^k U_i\right) - \left(\frac{1}{k}\sum_{i=1}^k \left(\frac{i}{k}\right)^{-\hat{\rho}} U_i\right)}{\left(\frac{1}{k}\sum_{i=1}^k \left(\frac{i}{k}\right)^{-\hat{\rho}}\right)\left(\frac{1}{k}\sum_{i=1}^k \left(\frac{i}{k}\right)^{-\hat{\rho}} U_i\right) - \left(\frac{1}{k}\sum_{i=1}^k \left(\frac{i}{k}\right)^{-2\hat{\rho}} U_i\right)}, \tag{2.24}$$

where U_i are the log-spacings $U_i := i(\log X_{n-i+1,n} - \log X_{n-i,n})$. For details on the derivation and the statistical behaviour of this estimator we refer to the paper. We will only use it and show how well it works for the data that we consider.

We still need to estimate the second order parameter ρ. This parameter is also important for the regression estimator. Different possibilities for estimating ρ are discussed in the next section.

2.3.4. Estimating the second order parameter

A method for estimating this ρ is proposed by Gomes and Martins (2002) and given by:

$$\hat{\rho}_\tau(k) = \hat{\rho}_n^{(\tau)}(k) := -\left|\frac{3(T_n^\tau(k)-1)}{T_n^\tau(k)-3}\right|, \tag{2.25}$$

where

$$T_n^\tau(k) := \begin{cases} \frac{\left(M_n^{(1)}(k)\right)^\tau - \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}}}{\left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} - \left(\frac{1}{6}M_n^{(3)}(k)\right)^{\frac{\tau}{3}}}, & \text{if } \tau > 0, \\ \frac{\log\left(M_n^{(1)}(k)\right) - \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right)}{\frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) - \frac{1}{3}\log\left(\frac{1}{6}M_n^{(3)}(k)\right)}, & \text{if } \tau = 0, \end{cases} \tag{2.26}$$

with $M_n^{(j)}$ given by (2.13). Here we see that the estimate still depends on a tuning parameter τ. Gomes and Martins (2002) suggest that for $\rho \in (-\infty,-1)$ we use $\tau = 0$, and for $\rho \in [-1,0)$ we use $\tau = 1$. The way we can implement this is by first choosing $\tau = 0$ and if this gives $\rho \in [-1,0)$ we will use this as the estimate and otherwise we set $\tau = 1$ and calculate ρ for this tuning parameter. If $\tau = 0$ gives $\rho \in [-1,0)$, we need to be sure that $\tau = 1$ gives $\rho \in [-1,0)$ too. This leads to the following claim.

Claim. *If $\tau = 0$ gives $\rho \in (-\infty,-1)$ then $\tau = 1$ gives $\rho \in (-\infty,-1)$ as well.*

Proof. Note that: $\rho < -1 \iff T_n^\tau > \frac{3}{2}$ or $T_n^\tau < 0$ from equation (2.25).
Assume $\tau = 0$ gives $\rho \in (-\infty,-1)$.
Then we distinguish two cases:

(1) $T_n^0 < 0$;

(2) $T_n^0 > \frac{3}{2}$.

If (1) holds, using equation (2.26), we can obtain that the following should hold too.

$$\log\left(M_n^{(1)}(k)\right) > \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) \text{ and } \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) < \frac{1}{3}\log\left(\frac{1}{6}M_n^{(3)}(k)\right) \text{ or}$$
$$\log\left(M_n^{(1)}(k)\right) < \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) \text{ and } \frac{1}{2}\log\left(\frac{1}{2}M_n^{(2)}(k)\right) > \frac{1}{3}\log\left(\frac{1}{6}M_n^{(3)}(k)\right).$$

From the monotonicity of the log it follows that we should also have that for $\tau > 0$:

$$\left(M_n^{(1)}(k)\right)^\tau > \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} \text{ and } \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} < \left(\frac{1}{6}M_n^{(3)}(k)\right)^{\frac{\tau}{3}} \text{ or}$$
$$\left(M_n^{(1)}(k)\right)^\tau < \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} \text{ and } \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{\tau}{2}} > \left(\frac{1}{6}M_n^{(3)}(k)\right)^{\frac{\tau}{3}}.$$

Thus $T_n^\tau(k) < 0$ as well for $\tau > 0$ and therefore we have that $\tau = 1$ gives $\rho \in (-\infty,-1)$.

If (2) holds this means that:

$$T_n^0(k) = \frac{\log a - \log b}{\log b - \log c} > \frac{3}{2},$$

for $a = M_n^{(1)}(k)$, $b = \left(\frac{1}{2}M_n^{(2)}(k)\right)^{\frac{1}{2}}$ and $c = \left(\frac{1}{6}M_n^{(3)}(k)\right)^{\frac{1}{3}}$. We now wish to show that $T_n^1(k) = \frac{a-b}{b-c} > \frac{3}{2}$ as well.

We will show that if $T_n^0(k) > 0$ then $T_n^1(k) > T_n^0(k)$, which suffices. Again we distinguish two cases:

(a) a > b > c;

(b) a < b < c.

Assume (a). Applying the mean value theorem, given in Stewart (2011), on $f(x) = \log x$ gives us that for some x and y for which $a > x > b > y > c$:

$$\frac{\log a - \log b}{a-b} = \frac{1}{x} < \frac{1}{y} < \frac{\log b - \log c}{b-c},$$

and therefore $T_n^1(k) = \frac{a-b}{b-c} > \frac{\log a - \log b}{\log b - \log c} = T_n^0(k) > \frac{3}{2}$, which is as desired.

Assuming (b) and again applying the mean value theorem gives the same result and therefore under (2) we obtain $T_n^1(k) > \frac{3}{2}$ implying $\rho \in (-\infty, -1)$ and under (1) we obtain $\rho \in (-\infty, -1)$ as well. Thus, when $\tau = 0$ gives $\rho \in (-\infty, -1)$, then $\tau = 1$ gives $\rho \in (-\infty, -1)$ as well. □

Gomes and Martins (2002) suggest to use the threshold $k_1 = \min\{n-1, \lfloor \frac{n}{\log\log n} \rfloor\}$, which leads to the following estimator of ρ:

$$\hat{\rho}_1 = \hat{\rho}_\tau(k_1). \tag{2.27}$$

Hall and Welsh (1985) introduced another estimator of ρ, which is given by:

$$\hat{\rho}_2 = \left| \log \left| \frac{\frac{1}{\hat{\gamma}^H(\lfloor n^{0.9} \rfloor)} - \frac{1}{\hat{\gamma}^H(\lfloor n^{0.5} \rfloor)}}{\frac{1}{\hat{\gamma}^H(\lfloor n^{0.95} \rfloor)} - \frac{1}{\hat{\gamma}^H(\lfloor n^{0.5} \rfloor)}} \right| \right| \cdot \left(\log n^{-0.05} \right)^{-1}. \tag{2.28}$$

We simply use these two estimators, compare their performance and check whether they are appropriate for the data. For the statistical behaviour of these estimators we refer to the papers that introduce the estimators.

For the regression estimator we can deal with ρ by taking a range of ρ's and for a threshold k fixed we perform the regression for all these ρ's. Then for each threshold we select ρ for which the squared errors of the residuals are minimized. We select the threshold after we have found this ρ and we write $\hat{\rho}^{MSE}$.

2.3.5. Pros and cons of the estimators

The Hill and Moment estimators are easiest to implement. For large tail indices $\gamma > 1$ the Hill estimator should work better than the Moment estimator, but the Hill estimator fails for $\gamma \leq 0$, which is the regime for which we need the Moment estimator or other alternatives. The other two and more technical estimators considered in this thesis are more difficult to implement, but should adjust for the bias in the Hill estimator and therefore be an improvement under some conditions. In Haan and Ferreira (2006) it is discussed that the tail index of log-return data is typically around 0.3, so we should still see if the improvements hold for data of $\gamma \approx 0.3$.

The adjusted Hill estimator assumes a parametric form given in equation (2.22) and based on this form it improves the estimate by adjusting the Hill estimator with parameters ρ and β. This parametric form does not necessarily hold for arbitrary financial data and it might be the case that the adjusted Hill estimator does not work for distributions where this assumption is not valid. In the simulations we will see how well this adjusted Hill estimator performs on samples where the assumption does and does not hold.

The Regression estimator does require that $\gamma > 0$ but does not require any parametric form and therefore seems to be more robust than the adjusted Hill estimator. Of all methods this is the most time consuming method, since we have to do the complete regression for every threshold.

2.4. Threshold selection for tail index estimators

In this section we introduce three methods for selecting the threshold k for the Hill estimator. The first method called stability method is a new method that we propose and the Guillou Hall method is obtained from the literature. Hall and Guillou (2001) discuss that the threshold can be chosen in the region where the stochastic fluctuations in γ are small for the first time. The idea behind this is that the variance in the estimator is probably small in this region because of the small fluctuations. Since we look at the first time this happens the bias should not be too large either. The precision of the estimator can therefore be expected to be high in this region. Based on this insight we choose the tail index in Figure 2.4. This method is intuitive and for a single Hill plot we can simply choose the threshold by looking at the plot. When we wish to do more tail index estimations, an automated procedure becomes more convenient. To be able to do this we need to mathematically quantify how we choose this threshold.

We propose the following. First we choose a domain of thresholds that is allowed, i.e. any interval for which $k \in \{k_{min}, \cdots, k_{max}\}$ where $k_{min} = \lfloor c_1 n^p \rfloor$ and $k_{max} = \lfloor c_2 n^p \rfloor$. Here $0 < c_1 < c_2$ and $0 < p < 1$ will make sure that $\frac{k}{n} \to \infty$ and that $k \to \infty$, which is what we need for the estimator to be asymptotically consistent. Also we should have a sufficient number of thresholds in the interval to be able to choose from and therefore p and $c_2 - c_1$ should not be too small. To determine the first stable period we define a quantity for the stability around k in equation (2.29).

$$S(k) = \sum_{i=k-m}^{k+m-1} \left(|\gamma(i+1) - \gamma(i)| \right), \tag{2.29}$$

which is a moving average of the absolute changes in the tail index. Here $m \in \mathbb{N}$ should be small such that this estimate is representative for k, but not too small, since we could be dealing with a coincidental stable period then. The smaller S the more stable the tail index is around k.
A possible threshold is defined to be:

$$k^S = \arg\min_k \{S(k) \cdot k^q\}, \tag{2.30}$$

where $q > 0$ and $S(k)$ defined above. This method is very easy to implement, gives freedom in the choice of q and is not time consuming. We can also use it on other estimators, since the bias-variance trade-off appears for other estimators as well. Suppose we have n data points $X_1, \cdots, X_n$, then we can compute the threshold for the Hill estimator by using the following algorithm.

Algorithm 2.2 (Stability method)**.**

---

for $k = m+1, \cdots, k_{max} - m$
  $S(k) := \sum_{i=k-m}^{k+m-1} \left(|\gamma^H(i+1) - \gamma^H(i)| \right)$
end for
$k^S := \arg\min_k \{S(k) \cdot k^q\}$
return k^S

Caeiro and Gomes (2016) introduce another method, which is also based on these stochastic fluctuations. The details of this method can be found in their paper and we will only summarize and use the result, given in Algorithm 2.3. Suppose we have n data points $X_1, \cdots, X_n$, then we can compute the threshold for the Hill estimator as follows.

Algorithm 2.3 (Guillou Hall method)**.**

Order the sample to get $\{X_{1,n}, \cdots, X_{n,n}\}$
$y_i := i\{\log X_{n-i+1,n} - \log X_{n-i,n}\}$
$T_n(k) := \sqrt{\frac{3}{k}} \frac{\sum_{i=1}^{k} (k-2i+1) y_i}{\sum_{i=1}^{k} y_i}$
$Q_n(k) := \sqrt{\frac{1}{\lfloor \frac{k}{2} \rfloor + 1} \sum_{j=k-\lfloor \frac{k}{2} \rfloor}^{k+\lfloor \frac{k}{2} \rfloor} T_n^2(j)}$
$k^{GH} := \inf\{k : Q_n(j) \geq c \ \forall j \geq k\}$
return k^{GH}

This threshold selection method works for the Hill estimator and appears to be a accurate threshold choice for $1.25 \leq c \leq 1.5$, see Caeiro and Gomes (2016). It makes sure that the variance and bias are small. Since the Moment estimator and the Adjusted Hill estimator are second order adjustments of the Hill estimator, we can expect those methods to have large stochastic fluctuations when the Hill estimator has large fluctuations. Therefore, this method might work on these estimators and we will apply this as well. The results will be given in the next section.

2.5. Results for the different thresholds methods and estimators

2.5.1. Simulation study

Now that we have methods to determine the threshold we can test these methods on simulated data. This test data is simulated from multiple distribution types with multiple known tail indices. The threshold method that has the smallest mean squared error (MSE) on the test data will be our preferred method. This method can be different for different sample sizes.

After we have chosen the preferred method for every sample size and every estimator that we considered we can compare these estimators on the test data. Assuming $\gamma > 0$, all the four estimators should give a reasonable estimate of γ. The performance, however, will depend on the tail index itself. We can determine which estimator or linear combination of estimators to use for the tail index estimation. Choosing the best linear combination is a large optimization problem. We will not consider all the possible linear combinations and also no initial guess will be used. We can already improve our estimate by simply considering the averages of all possible combinations.

The results of the threshold selection methods on the four estimators are given in the next section, where we choose the threshold selection method with the minimum mean squared error (minimum MSE). After choosing a method for each estimator we can consider averages of combinations and we will show that this improves our estimate with respect to MSE.

Since we wish to use the estimates on financial data, we should test the estimators on data with a tail index in the same range as tail indices from financial data. In Haan and Ferreira (2006) it is discussed that the tail index of financial data is typically around 0.3 such that the variance exists, but the skewness and kurtosis do not necessarily exist. To test the estimators and threshold selection methods we test the methods on samples from the Pareto, Burr and Fréchet distribution, all with tail indices ranging from 0.1 to 0.5. First we test the threshold methods on the estimators. Afterwards we use these methods on the estimators and see which estimator works best for which sample.

2.5.2. Threshold methods Hill estimator

We first analyse the threshold methods for the Hill estimator. We also use the stability method, that is given in Algorithm 2.2, with $c_1 = 0.2$, $c_2 = 3$, $p = \frac{1}{2}$ and $q = 0, 0.5, 1, 1.5, 2, 3$. Considering additional values of q might improve the result slightly, but we will see that the estimate is not strongly dependent on q for q values around the optimal value. Lastly we use the Guillou and Hall method with $c = 1.25, 1.5$, which is given in Algorithm 2.3. The methods are tested on data generated from Pareto, Burr, and Fréchet distributions with sample sizes of 300, 1000 and 10000. The performance of the estimate depends on the underlying distribution and on the sample size. For a Fréchet and Burr distribution all the threshold methods work quite well since the difference between the mean of the estimates is small indicating a small bias, but the MSEs differ significantly for different methods. We cannot safely say that one method is the best method, since no method is best for all underlying distributions. To select our method we compute the average MSE over all the different underlying distributions (AMSE) for every threshold selection method and use the one with the smallest AMSE value.

The results for the AMSE of each threshold selection method are given in Table 2.2. The stability method for $q \geq 1.5$ is never among the best in performance. The stability methods with $q = 0, 0.5, 1$ perform well compared to the others for all data sizes and the results are quite similar for the three values. The Guillou Hall method performs well for the largest data set. Since the stability method with $q = 0.5$ performs best for sample sizes of 1000 and 10000 and second best for size 300 we choose to work with this stability method for the Hill estimator independent of the sample size throughout the rest of this thesis.

Table 2.2: **Average mean squared errors of the Hill estimates.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the stability and Guillou Hall threshold selection methods for sample sizes of 300, 1000 and 10000.

		Sample size		
Method	Parameter	300	1000	10000
Stability	q=0	† 0.127	0.102	0.065
Stability	q=0.5	0.128	† 0.100	† 0.063
Stability	q=1	0.131	0.100	0.066
Stability	q=1.5	0.138	0.118	0.105
Stability	q=2	0.157	0.148	0.142
Stability	q=3	0.176	0.174	0.170
Guillou Hall	c=1.25	0.173	0.118	0.063
Guillou Hall	c=1.5	0.189	0.134	0.069

†Denotes the best estimator for this sample size.

2.5.3. Threshold methods Moment estimator

When we work with the Moment estimator we cannot simply apply all the methods from the previous sections. The stability estimator should however still work, because the first stable period for the Moment estimator is also the threshold range where the variance has decreased and the bias is still relatively small. For the same reason we can use the Guillou and Hall method.

Since the Moment estimator corrects for the Hill bias we can expect the bias to be smaller. Therefore we will have the same bias for a larger threshold which is why we will consider a larger range of thresholds. As a consequence we will choose $c_2 = 5$ and $c_1 = 0.2$ and leave the other parameters unchanged with respect to previous section. The resulting AMSEs are given in Table 2.3.

Table 2.3: **Average mean squared errors of the Moment estimates.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the stability and Guillou Hall threshold selection methods for sample sizes of 300, 1000 and 10000.

		Sample size		
Method	Parameter	300	1000	10000
Stability	q=0	† 0.128	0.092	0.053
Stability	q=0.5	0.130	0.097	0.061
Stability	q=1	0.139	0.111	0.089
Stability	q=1.5	0.156	0.145	0.135
Stability	q=2	0.192	0.183	0.180
Stability	q=3	0.232	0.223	0.224
Guillou Hall	c=1.25	0.164	0.095	0.035
Guillou Hall	c=1.5	0.145	† 0.080	† 0.032

†Denotes the best estimator for this sample size.

Here we see that the Guillou Hall method with $c = 1.5$ performs best for sample sizes of 1000 and 10000, but for size 300 the stability method with $q = 0$ is best and clearly better than the Guillou Hall method. Depending on the sample size we choose one of these methods throughout the rest of this thesis. With this choice, for sample sizes of 300 the AMSE is comparable to the Hill estimator and for the other sample sizes it performs better.

2.5.4. Threshold methods Regression estimator

For the regression estimator we have to choose a method for the estimation of ρ as well as a method for the determination of the threshold. We run the regression on the same samples as in the previous sections for first $\rho = -1$ and then for ρ estimated by equation (2.27). We do not consider all the possible estimators,

since according to Vandewalle et al. (2004), the regression is not very much dependent on ρ. After these estimations we determine the threshold for each of these choices of ρ. We do this using the stability method, giving k^S, and by choosing the threshold that minimizes the mean squared error multiplied by the number of explanatory variables (i.e. the threshold), giving k^{MSE}. We consider the same threshold range as with the Moment estimator, since the Regression estimator should also have a lower bias than the Hill estimator. Lastly we let $\rho = -2, -1.9, \cdots, -0.1$ and choose the ρ_{MSE} and k^{MSE} simultaneously. The summarized results can be found in Table 2.4.

Table 2.4: **Average mean squared errors of the Regression estimates.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the stability and Guillou Hall threshold selection methods for sample sizes of 300, 1000 and 10000.

			Sample size		
Method	Parameter	$\hat{\rho}$	300	1000	10000
Minimizing the error		$\hat{\rho}_{MSE}$	0.184	0.161	0.142
Minimizing the error		$\hat{\rho}_1$	1.578	0.826	0.172
Minimizing the error		$\hat{\rho} = -1$	0.235	0.198	0.177
Stability	$q = 0$	$\hat{\rho}_1$	0.495	0.379	† 0.056
Stability	$q = 0$	$\hat{\rho} = -1$	0.425	0.382	0.058
Stability	$q = 0.5$	$\hat{\rho}_1$	0.518	0.361	0.085
Stability	$q = 0.5$	$\hat{\rho} = -1$	† 0.133	† 0.097	0.057
Stability	$q = 1$	$\hat{\rho}_1$	0.144	0.106	0.060
Stability	$q = 1$	$\hat{\rho} = -1$	0.166	0.135	0.090

†Denotes the best estimator for this sample size.

For sample sizes of 300 and 1000 we see that the best combination is the stability method with $q = 0.5$ and $\hat{\rho} = -1$. For a sample size of 10000 the performance of this method is comparable to the performance stability method with $q = 0$ and $\hat{\rho} = \hat{\rho}_1$. For convenience we use the method with $q = 0.5$ and $\hat{\rho} = -1$ for all sample sizes throughout the rest of this thesis, even though it performs second best on the test data for size 10000.

2.5.5. Threshold methods Adjusted Hill estimator

The Adjusted Hill estimator uses the Hill estimator with an adjustment dependent on second order parameters ρ and β. Besides those two parameters, we also have to choose a threshold. We used the stability method and the Guillou Hall method for this threshold selection. ρ is estimated by equation (2.27) and (2.28) and β by equation (2.24). We choose the ranges of thresholds considered to be the same as with the Hill estimator and the Moment estimator and choose the range which performs best.

Table 2.5: **Average mean squared errors of the Adjusted Hill estimates.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the stability and Guillou Hall threshold selection methods for sample sizes of 300, 1000 and 10000.

			Sample size		
Method	Parameter	$\hat{\rho}$	300	1000	10000
Stability	$q = 0$	$\hat{\rho}_1$	0.140	0.102	0.065
Stability	$q = 0$	$\hat{\rho}_2$	0.134	0.095	0.059
Stability	$q = 0.5$	$\hat{\rho}_1$	0.141	0.100	0.063
Stability	$q = 0.5$	$\hat{\rho}_2$	† 0.134	† 0.095	†0.059
Stability	$q = 1$	$\hat{\rho}_1$	0.140	0.102	0.065
Stability	$q = 1$	$\hat{\rho}_2$	0.135	0.099	0.065
Guillou Hall	$c = 1.25$	$\hat{\rho}_1$	0.675	0.153	0.067
Guillou Hall	$c = 1.25$	$\hat{\rho}_2$	0.150	0.124	0.064
Guillou Hall	$c = 1.5$	$\hat{\rho}_1$	0.301	0.125	0.066
Guillou Hall	$c = 1.5$	$\hat{\rho}_2$	0.316	0.101	0.063

† Denotes the best estimator for this sample size.

Here the stability method with $q = 0.5$ and $\hat{\rho} = \hat{\rho}_2$ performs best for all sample sizes (the difference with the stability method with $q = 0$, is made only in the fourth decimal). Therefore, when using the Adjusted Hill estimator, we will work with the stability method with $q = 0.5$ throughout the rest of this thesis.

2.5.6. Average of estimators

All the four estimators give a reasonable estimate of γ. The performance, however, depends on the tail index itself as we can see from the overview in the appendix. Considering a weighted average of these four estimators should also give a reasonable estimate and by using a model averaging approach one can determine those weights. Model averaging is a large field of study and beyond the scope of this thesis. For determining our final estimate we will only consider the four estimators and the averages of all possible combinations of two, three and four estimators.

In Tables B.1, B.2 and B.3 in the appendix the average mean squared errors are given of all these estimates for sample sizes 300, 1000 and 10000, respectively. Estimating γ as the average of the Hill, Moment and Regression estimator gives the best results for sample sizes 300 and 1000. For a sample size of 10000 the average of the Moment, Regression and Adjusted Hill estimator performs best. The final estimator we will use in the rest of this thesis will therefore be this average of three of the four estimtors. The results of this estimator compared with the individual estimators are given Tables 2.6, 2.7 and 2.8.

In conclusion, for every estimator, except the Moment estimator we can choose the threshold selection method independent of the sample size. For this Moment estimator we are better off choosing a different method for the sample size of 300, namely the stability method with $q = 0$ instead of the Guillou Hall method with $c = 1.5$. The stability method with $q = 0.5$ performs best (or almost best) for all other estimators. The best choice for the second order parameter is different for the Adjusted Hill estimation and the Regression estimation. In the Regression method we choose $\hat{\rho} = -1$ and in the Guillou Hall method we choose $\hat{\rho} = \hat{\rho}_2$. With those methods we calculate four estimates and depending on the sample size we take the average of the Moment, Regression and Hill or Adjusted Hill estimator. The final results of these methods are given in Tables 2.6, 2.7 and 2.8

Table 2.6: **Performance in AMSE of the final estimate for sample size** 300**.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the final estimator for a sample size of 300. For the same samples we also show the AMSE of the Hill, Moment, Regression and Adjusted Hill estimator.

	γ					
Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill	0.137	0.126	0.122	0.130	0.137	0.131
Moment	0.112	0.134	0.141	0.145	0.147	0.136
Regression	0.132	0.132	0.129	0.149	0.165	0.142
Adjusted Hill	0.133	0.126	0.124	0.136	0.150	0.134
Final estimate	0.116	0.117	0.119	0.128	0.136	0.124

Table 2.7: **Performance in AMSE of the final estimate for sample size** 1000**.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the final estimator for a sample size of 1000. For the same samples we also show the AMSE of the Hill, Moment, Regression and Adjusted Hill estimator.

	γ					
Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill	0.117	0.099	0.090	0.089	0.098	0.099
Moment	0.084	0.097	0.105	0.102	0.104	0.099
Regression	0.106	0.094	0.095	0.101	0.118	0.103
Adjusted Hill	0.107	0.091	0.087	0.090	0.099	0.095
Final estimate	0.089	0.088	0.085	0.091	0.098	0.090

Table 2.8: **Performance in AMSE of the final estimate for sample size** 10000**.**

This table shows the square root of the average mean squared error ($\sqrt{\text{AMSE}}$) for the final estimator for a sample size of 10000. For the same samples we also show the AMSE of the Hill, Moment, Regression and Adjusted Hill estimator.

	γ					
Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill	0.088	0.067	0.058	0.054	0.052	0.065
Moment	0.058	0.056	0.062	0.056	0.057	0.058
Regression	0.075	0.059	0.050	0.054	0.057	0.060
Adjusted Hill	0.077	0.061	0.053	0.053	0.053	0.060
Final estimate	0.060	0.050	0.048	0.051	0.052	0.052

As expected, the final estimate is better than all the other estimators on average for the three sample sizes and the 15 distributions considered. We should here realize that average of three is however not best for all γ. It is for example outperformed by the Moment estimator for $\gamma = 0.1$, but since we do not know γ a priori and since Haan and Ferreira (2006) discusses that γ is typically around 0.3, we will use the final estimate (i.e. linear combination of three estimators) as the estimate for γ.

2.6. Conclusion

From the literature we know that log-return type financial variables typically show heavy-tailed behaviour in the lower tail. Under this assumption we can use the estimators explained in this section, which work for $\gamma > 0$. For all estimators we need to choose a threshold and we found threshold selection methods from the literature for each estimator, except for the Regression estimator. Furthermore, we came up with a new

threshold selection method, the stability method, which is based on the bias-variance trade-off we have to make in tail index estimations. This stability method performs better than the considered methods from the literature for the adjusted Hill and Hill estimator, but not for the Moment estimator. We also used it for the Regression estimator to determine the threshold there.

The MSEs of all estimators are close over the considered range of tail indices. Some estimators perform better for smaller tail indices and some for larger tail indices. By combining the estimators and taking the average over three of the four we improve the estimators MSE. The choice here depends on the sample size. Thus, we now have an automated procedure for estimating tail indices of heavy-tailed data for different sample sizes. We can model the tails of the distributions of these heavy-tailed variables more accurately. Using estimates we can now simulate financial variables with similar tails to what has been observed historically. Furthermore, we can obtain appropriate estimates of risk measures for tail risk such as the VaR_p for low p.

3

VAR models and tail indices

3.1. Introduction

In finance and economics data often takes the form of a time series. A simple time series model that can be used to describe a single financial time series is the autoregressive (AR) model. Here the time series variable depends on previous values, a constant and an error term, usually assumed to be white noise. Another model that can be used is a moving average (MA) model, where the value of the variable is a linear combination of error terms. Both of these models will be introduced and elaborated on in the next section. Subsequently, we will extend these models to the multivariate case, which will give us a vector autoregressive (VAR) and vector moving average (VMA) model, respectively.

At Ortec Finance, a large number of economic variables are decomposed into three frequency components. Every frequency component is then modelled as a function of factors and an error term. These factors capture the complete dynamics of a frequency band of the economy as a whole and are described by a VAR model. The time series variables we are dealing with are possibly heavy-tailed. A kernelling method is used in conjunction with a moment-matching method to account for these heavy tails. Here the moment-matching is used to adjust the distribution function obtained from kernelling, such that the third and fourth moment match the third and fourth moment of the data. We wish to apply extreme value theory here to model the tails of the distribution and explore how this can be implemented in the current approach of Ortec. We will show analytically that, under some conditions, the tail index of the errors equals the tail index of the time series variable and we will show what this means for the tail index estimation. This also shows that for a time series that is heavy-tailed, we should model the error as a heavy-tailed random variable. Definitions of this section and more details on this topic can be found in Shumway and Stoffer (2010).

3.2. AR and MA models

The definition of an autoregressive model, which can be used to describe certain one-dimensional time series in finance, is given in Definition 3.1.

Definition 3.1 (AR model)**.** *An autoregressive model of order p (AR(p) model) is of the form:*

$$y_t = c + \sum_{i=1}^{p} \phi_p y_{t-p} + e_t, \tag{3.1}$$

where c and ϕ_i are constants $\forall i$, with $\phi_p \neq 0$ and e_t is zero-mean white noise.

We can assume $c = 0$ for any AR(p) model, since we can de-mean the series y_t to obtain:

$$x_t = y_t - \mu = \sum_{i=1}^{p} \phi_p (y_{t-p} - \mu) + e_t = \sum_{i=1}^{p} \phi_p x_{t-p} + e_t, \tag{3.2}$$

where $\mu = c\left(1 - \sum_{i=1}^{p} \phi_p\right)^{-1}$ and proceed with the analysis of x_t, which can in the end easily be transformed back to y_t. We also introduce a more compact form, where we use the so-called autoregressive operator.

Definition 3.2. *We define the autoregressive operator as:*

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \tag{3.3}$$

where the backshift operator B is such that $Bx_t = x_{t-1}$.

Then we can rewrite equation (3.1) as:

$$\phi(B) y_t = e_t, \tag{3.4}$$

where the ϕ-function completely captures the autoregressive part. A time series can also be modelled as an MA model, which is given by the definition below.

Definition 3.3 (MA model)**.** *A moving average model of order q (MA(q) model) is of the form:*

$$y_t = c + \sum_{i=0}^{q-1} \theta_i e_{t-i}, \tag{3.5}$$

where $c, \theta_i \in \mathbb{R}\ \forall i$ are constants, $\theta_q \neq 0$ and e_t is zero-mean white noise.

Important characteristics of time series are stability and stationarity, which we will give in two forms, namely strongly stationary and weak-sense stationary. Both of these characterestics are introduced below.

Definition 3.4. *We say that a process is strongly stationary if $F(y_t) = F(y_s)\ \forall t, s$, with $F(y)$ the distribution function of y (i.e. the probability distribution is not a function of time).*

Definition 3.5. *We say that a process is stationary in a weak sense if $\mathbb{E}[x_t] = \mathbb{E}[x_s]\ \forall t, s$ and $\mathrm{Cov}(x_t, x_s) = \mathbb{E}[(x_s - \mathbb{E}[x_s])(x_t - \mathbb{E}[x_t])] = \mathbb{E}[(x_{|t-s|} - \mathbb{E}[x_{|t-s|}])(x_0 - \mathbb{E}[x_0])] = \mathrm{Cov}(x_{|t-s|}, x_0)\ \forall t, s$ (i.e. the mean does not vary with respect to time and the autocovariance is a function of the lag $|t-s|$ only).*

From now on we will use the term stationary to indicate stationary in a weak sense. It follows directly that strong stationarity implies stationarity if the variance is finite. $F(y_t) = F(y_s)$ namely implies $f(y_t) = f(y_s)$ almost surely, which implies $\mathbb{E}[x_t] = \mathbb{E}[x_s]$ and $\mathbb{E}[(x_s - \mathbb{E}[x_s])(x_t - \mathbb{E}[x_t])] = \mathbb{E}[(x_{|t-s|} - \mathbb{E}[x_{|t-s|}])(x_0 - \mathbb{E}[x_0])]$. The reverse is not necessarily true, see the following example.

Example 3.1. *Consider x_t to be normally distributed white noise for $t < 1$ and t-distributed white noise for $t \geq 1$, such that the variances equal $\sigma\ \forall t$.*
Then x_t is not strongly stationary, since the distribution function is not constant in t. It is however weak-sense stationary, since the autocovariance equals σ for $t = s$ (i.e. for $|t-s| = 0$) and 0 elsewhere. Therefore the autocovariance function is a function of the lag $|t-s|$ only. Furthermore, the mean is 0 everywhere and thus independent of t.

In financial data we often work with the log-differences of time series, for example the log-return of a stock. This is the log-difference of the stock price. These log-returns are usually assumed to be stable, meaning that the errors in an early stage do not blow up at a later stage.

Definition 3.6. *An AR(p) process given by equation (3.1) is stable if and only if the roots of the lag polynomial lie outside the unit circle (i.e. solutions of $0 = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ have that $|z| < 1$).*

Theorem 3.1. *An AR(p) process which is stable and has finite variance errors is stationary.*

The proof of this theorem is given in Shumway and Stoffer (2010). They also state that a time series described by an AR(1) model can be written as the weighted sum of a countably infinite number of errors of the model (i.e. an MA(∞) model) if this AR model is stationary, which leads to the following proposition.

Proposition 3.2. *Any stationary AR(1) model can be written as a stationary MA(∞) model.*

Proof. Let y_t be described by a stationary AR(1) model, meaning that we can write $y_t = \phi y_{t-1} + e_t$ with $0 < |\phi| < 1$, where e_t is zero-mean white noise. Then we can rewrite as follows:

$$y_t = \phi y_{t-1} + e_t = \phi(\phi y_{t-2} + e_{t-1}) + e_t = \phi^2 y_{t-2} + \phi e_{t-1} + e_t = \cdots = \sum_{i=0}^{\infty} \phi^i e_{t-i},$$

which is an MA(∞) model with $\theta_i = \phi^i$. □

In the next section, we will generalize this result for vector autoregressive models (VAR model) with an arbitrary lag p.

3.3. VAR models

To gain insight into the dynamics of a number of financial time series, we can extend the autoregressive models to the multivariate case. This extension gives us a vector autoregressive (VAR) model, which is formally defined below. In this model a vector of economic variables depends linearly on previous vectors (via a matrix vector product) and an error term. Due to the matrix product, different time series can exhibit correlation even though the error terms can be uncorrelated.

Definition 3.7 (VAR model)**.** *A vector autoregressive model of order p (VAR(p) model) is of the form:*

$$y^{(t)} = c + \sum_{i=1}^{p} A_i y^{(t-i)} + \epsilon^{(t)}, \tag{3.6}$$

where $c \in \mathbb{R}^n$ and $A_i \in \mathbb{R}^{n\times n}$ are constants $\forall i$, A_p has at least one non-zero element and $y^{(t)}, \epsilon^{(t)} \in \mathbb{R}^n$, with $\mathbb{E}[\epsilon^{(t)}] = 0$, $\mathbb{E}[\epsilon^{(t)}\epsilon^{(t)\prime}] = \Omega$ the contemporaneous covariance matrix and $\mathbb{E}[\epsilon^{(t)}\epsilon^{(t-k)\prime}] = 0\ \forall k > 0$.

Similar to the AR(p) model, we can assume $c = 0$ without loss of generality.

Definition 3.8 (VMA model)**.** *A vector moving average model of order q (VMA(q) model) is of the form:*

$$y^{(t)} = c + \sum_{i=0}^{q} A_i \epsilon^{(t-i)}, \tag{3.7}$$

where $c \in \mathbb{R}^n$ and $A_i \in \mathbb{R}^{n\times n}$ are constants $\forall i$, A_q has at least one non-zero element and $\epsilon^{(t)}$ is zero-mean white noise.

We wish to express the resulting tail index of a VAR(p) time series in terms of the error tail indices. To be able to do this we will rewrite the VAR(p) model.

Theorem 3.3. *Any stationary n-dimensional VAR(p) model can be written as a stationary $n \cdot p$-dimensional VMA(∞) model.*

The proof of this theorem follows directly from the two lemmas below, of which we will give the proofs.

Lemma 3.4. *Any stationary n-dimensional VAR(p) model can be written as a stationary $n \cdot p$-dimensional VAR(1) model.*

Proof. Let the n-dimensional VAR(p) model be given by equation (3.6), such that we have

$$y^{(t)} = A_1 y^{(t-1)} + A_2 y^{(t-2)} + \cdots + A_p y^{(t-p)} + \epsilon^{(t)}.$$

Now choosing $A \in \mathbb{R}^{n\cdot p \times n\cdot p}$ and $e^{(t)} \in \mathbb{R}^{n\cdot p}$ as follows:

$$A = \begin{pmatrix} A_1 & A_2 & A_3 & \cdots & A_p \\ I & 0 & 0 & \cdots & 0 \\ 0 & I & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & I & 0 \end{pmatrix} \quad \text{and} \quad e^{(t)} = \begin{pmatrix} \epsilon^{(t)} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \ \forall t,$$

gives us that

$$\tilde{y}^{(t)} = A\tilde{y}^{(t-1)} + e^{(t)} \tag{3.8}$$

where $\tilde{y}^{(t)} = \left(\left(y^{(t)}\right)^T \left(y^{(t)}\right)^T \cdots \left(y^{(t-p+1)}\right)^T\right)^T$. Then the vector $\tilde{y}^{(t)} \in \mathbb{R}^{n\cdot p}$ follows a VAR(1) and the first n components correspond to the components of y^t, which was the original time series from the n-dimensional VAR(p) model. □

Lemma 3.5. *Any stationary VAR(1) model can be written as a stationary VMA(∞) model.*

Proof. Let a VAR(1) model be given by

$$y^{(t)} = Ay^{(t-1)} + e^{(t)} \tag{3.9}$$

Then we can rewrite

$$y^{(t)} = A(Ay^{(t-2)} + e^{(t-1)}) + e^{(t)} = A^2 y^{(t-2)} + Ae^{(t-1)} + e^{(t)} = \cdots = \sum_{j=0}^{\infty} A^j e^{(t-j)},$$

which is the representation of a VMA(∞) model. □

Combining the results of Lemma 3.4 and Lemma 3.5 proves Theorem 3.3.

Theorem 3.6. *All components of $y^{(t)}$, which follows a VAR(p) process, can be written as a countably infinite linear combination of the errors of the VAR(p) process.*

Proof. Suppose we have an n-dimensional VAR(p) model defined by equation (3.6), then for arbitrary component i of $y^{(t)}$ we can write:

$$\begin{aligned}
y_i^{(t)} &= \sum_{j=0}^{\infty} \left(A^j e^{(t-j)}\right)_i \\
&= \sum_{j=0}^{\infty} \left(A^j\right)_i e^{(t-j)} \\
&= \sum_{j=0}^{\infty} \sum_{k=1}^{n\cdot p} \left(A^j\right)_{i,k} e_k^{(t-j)} \\
&= \sum_{j=0}^{\infty} \sum_{k=1}^{n\cdot p} \left(A^j\right)_{i,k} e_k^{(t-j)} \\
&= \sum_{j=0}^{\infty} \sum_{k=1}^{n} \left(A^j\right)_{i,k} \epsilon_k^{(t-j)},
\end{aligned}$$

where $\left(A^j\right)_i$ is the i'th row of A^j and then $(A^j)_{i,k}$ is simply the matrix element of A^j at row i and column k. All the $\epsilon_k^{(t-j)}$'s are errors with tail index γ_k. Since this is a countable number of variables and $\epsilon^{(t)}$ has the same underlying distribution for every t, we can rewrite this even further and see that $\tilde{y}_i^{(t)}$ can be written as follows:

$$\tilde{y}_i^{(t)} = \sum_{m=0}^{\infty} c_m \epsilon_{k_m}^{(t_m)}, \tag{3.10}$$

where $\epsilon_{k_m}^{(t_m)}$ follows a distribution with tail index γ_{k_m} for all m. □

The concepts of stationarity and stability can be extended from the one-dimensional case to the multi-dimensional case and we obtain the following.

Definition 3.9. *$y^{(t)}$ described by a VAR(p) model is weak-sense (or strongly) stationary if and only if each component of $y_i^{(t)}$ is weak-sense (or strongly) stationary.*

Definition 3.10. *Let $y^{(t)}$ be described by a VAR(1) model given by equation (3.9). $y^{(t)}$ is said to be stable if and only if all the eigenvalues of A lie within the unit circle.*
Note: A VAR(p) process is stable if and only if its VAR(1) representation is stable, see Shumway and Stoffer (2010).

Theorem 3.7. *A VAR(p) process which is stable and has errors with finite variances for all components is stationary.*

When the errors are described by a heavy-tailed distribution, we are interested in the tail index of the time series. Since we can write any time series described by a VAR(p) model as a VMA(∞) model, calculating the tail index of the VAR(p) model is equivalent to calculating the tail index of the corresponding VMA(∞) model. Every single component of this time series can then be described by a countably infinite weighted sum of random variables. Therefore, calculating the tail index of a time series described by a VAR model corresponds to calculating the tail index of a countably infinite weighted sum of random variables with different tail indices. The latter calculation will be the topic of the next section.

3.4. Linear combinations of independent random heavy-tailed variables

As a start we will calculate the tail index of a finite weighted sum of heavy-tailed variables. First we will show what happens when we scale a variable and then we will show what happens when we take a sum of heavy-tailed variables. We combine these to get the tail of a weighted sum and by taking limits we will obtain the final result for a countably infinite weighted sum. These three steps will be the topic of this section and in section 3.5 we will show what this means for arbitrary VAR models.

3.4.1. Scaling a heavy-tailed random variable

For a random variable X from a heavy-tailed distribution with upper tail index γ and lower tail index γ^L we have that for $x \to \infty$:

$$\mathbb{P}(X > x) \to Ax^{-\frac{1}{\gamma}} + \mathcal{O}\left(x^{-\beta}\right) \quad \text{and} \quad \mathbb{P}(X < -x) \to A^L x^{-\frac{1}{\gamma_L}} + \mathcal{O}\left(x^{-\beta^L}\right)$$

Scaling X by parameter $c > 0$ does not change the tail index, since:

$$\mathbb{P}(cX > x) = \mathbb{P}\left(X > \frac{x}{c}\right) \to A\left(\frac{x}{c}\right)^{-\frac{1}{\gamma}} + \mathcal{O}\left(x^{-\beta}\right) = \tilde{A}x^{-\frac{1}{\gamma}} + \mathcal{O}\left(x^{-\beta}\right).$$

Scaling X by parameter $-b = c < 0$ interchanges the tail indices, because then:

$$\mathbb{P}(cX > x) = \mathbb{P}\left(X < \frac{x}{c}\right) = \mathbb{P}\left(X < -\frac{x}{b}\right) \to A^L\left(\frac{x}{c}\right)^{-\alpha^L} + \mathcal{O}\left(x^{-\beta^L}\right) = \tilde{B}x^{-\alpha^L} + \mathcal{O}\left(x^{-\beta^L}\right) \text{ and similarly}$$

$$\mathbb{P}(cX < -x) \to \tilde{C}x^{-\alpha} + \mathcal{O}\left(x^{-\beta}\right)$$

To see what happens in the tail when we take the sum of two heavy-tailed distributions, we need some more calculations. We do this in the next section.

3.4.2. Sum of heavy-tailed random variables

Theorem 3.8. *Suppose we have X_1 and X_2 independent from a heavy-tailed distribution with upper tail indices $\gamma_1 = \frac{1}{\alpha_1}$ and $\gamma_2 = \frac{1}{\alpha_2}$, respectively, where we assume $\gamma_1 \neq \gamma_2$. Also suppose that the density functions f_1 and f_2 are bounded almost surely by some $C_1 \in \mathbb{R}$ and $C_2 \in \mathbb{R}$. Then the tail index of $X_1 + X_2$ is given by* $\max\{\gamma_1, \gamma_2\}$ *(the largest tail index survives in the upper tail).*

Proof. Without loss of generality assume $0 < \alpha_1 < \alpha_2$. We will show that for large s: $\mathbb{P}(X_1 + X_2 > s) \approx Cs^{-\alpha_1}$ for some constant C and therefore the tail index of the sum is the tail index of the heaviest individual tail.
$\mathbb{P}(X_1 + X_2 > s)$ can be calculated by making a partition of the complement set $\{X_1 + X_2 \leq s\}$ and calculating the probabilities of X_1 and X_2 to be in this partition. The idea comes from Hyung and de Vries (2007), but here we relax the condition $2 < \alpha_1 < \alpha_2$ to the more general case $0 < \alpha_1 < \alpha_2$. First we create a partition such that $\{X_1 + X_2 \leq s\} = A \cup B \cup C \cup D \cup E$, where:

$$\begin{aligned}
A &= \left\{X_1 + X_2 \leq s, X_1 > \frac{-s}{2}, X_2 > \frac{s}{2}\right\}, \\
B &= \left\{X_1 + X_2 \leq s, X_1 > \frac{s}{2}, X_2 > \frac{-s}{2}\right\}, \\
C &= \left\{X_1 \leq \frac{s}{2}, X_2 \leq \frac{s}{2}\right\}, \\
D &= \left\{X_1 + X_2 \leq s, X_1 \leq \frac{-s}{2}, X_2 > \frac{s}{2}\right\}, \\
E &= \left\{X_1 + X_2 \leq s, X_1 > \frac{s}{2}, X_2 \leq \frac{-s}{2}\right\}.
\end{aligned} \tag{3.11}$$

Then we have that $\mathbb{P}(X_1 + X_2 > s) = 1 - \mathbb{P}(A) - \mathbb{P}(B) - \mathbb{P}(C) - \mathbb{P}(D) - \mathbb{P}(E)$ and we can proceed by calculating these probabilities separately. $\mathbb{P}(C)$ is the easiest and for large s we have:

$$\begin{aligned}
\mathbb{P}(C) &= \mathbb{P}\left(X_1 \leq \frac{s}{2}, X_2 \leq \frac{s}{2}\right) \\
&= \mathbb{P}\left(X_1 \leq \frac{s}{2}\right)\mathbb{P}\left(X_2 \leq \frac{s}{2}\right) \\
&= \left(1 - A_1\left(\frac{s}{2}\right)^{-\alpha_1} + \mathcal{O}\left(s^{-\beta}\right)\right)\left(1 - A_2\left(\frac{s}{2}\right)^{-\alpha_2} + \mathcal{O}\left(s^{-\beta}\right)\right) \\
&= 1 - A_1 2^{\alpha_1} s^{-\alpha_1} + \mathcal{O}\left(s^{-\beta}\right),
\end{aligned}$$

where $\beta > \alpha_1$. $\mathbb{P}(A)$ requires more computation, which is given below:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(X_1 + X_2 \le s, X_1 > \frac{-s}{2}, X_2 > \frac{s}{2}\right)\\ &= \int_{\frac{-s}{2}}^{\frac{s}{2}} \left[F_2(s-x) - F_2\left(\frac{s}{2}\right)\right] f_1(x)dx\\ &= \int_{\frac{-s}{2}}^{\frac{s}{2}} F_2(s-x) f_1(x)dx - \int_{\frac{-s}{2}}^{\frac{s}{2}} F_2\left(\frac{s}{2}\right) f_1(x)dx\\ &= \mathrm{I}_A - \mathrm{II}_A.\end{aligned}$$

Here I_A can be worked out using a Taylor expansion for $(s-x)^{-\alpha_1}$ and $(s-x)^{-\beta}$. We can use the fact that for $x \in \left(-\frac{s}{2}, -\frac{s}{2}\right)$ we have $(s-x)^{-y} = s^{-y} + yxs^{-y-1} + \frac{(y+1)y}{2}x^2(s-q_y)^{-y-2}$ for some $q_y \in \left(-\frac{s}{2}, \frac{s}{2}\right)$, see Stewart (2011). From this, I_A can be calculated as follows:

$$\begin{aligned}\mathrm{I}_A &= \int_{\frac{-s}{2}}^{\frac{s}{2}} F_2(s-x) f_1(x)dx\\ &= \int_{\frac{-s}{2}}^{\frac{s}{2}} \left(1 - A_2(s-x)^{-\alpha_2} + \mathcal{O}\left((s-x)^{-\beta_2}\right)\right) f_1(x)dx\\ &= \int_{\frac{-s}{2}}^{\frac{s}{2}} \left(1 - A_2 s^{-\alpha_2} + \mathcal{O}\left(s^{-\beta_2}\right)\right) f_1(x)dx - \int_{\frac{-s}{2}}^{\frac{s}{2}} \left(A_2\alpha_2 s^{-\alpha_2-1} + \mathcal{O}\left(s^{-\beta_2-1}\right)\right) x f_1(x)dx\\ &\qquad - \int_{\frac{-s}{2}}^{\frac{s}{2}} \left(A_2 \frac{(\alpha_2+1)\alpha_2}{2}(s-q_{\alpha_2})^{-\alpha_2-2} + \mathcal{O}\left((s-q_{\beta_2})^{-\beta_2-2}\right)\right) x^2 f_1(x)dx\\ &= a\int_{\frac{-s}{2}}^{\frac{s}{2}} f_1(x)dx - b\int_{\frac{-s}{2}}^{\frac{s}{2}} x f_1(x)dx - c\int_{\frac{-s}{2}}^{\frac{s}{2}} x^2 f_1(x)dx,\end{aligned}$$

where

$$\begin{aligned}a &= \left(1 - A_2 s^{-\alpha_2} + \mathcal{O}\left(s^{-\beta_2}\right)\right),\\ b &= \left(A_2\alpha_2 s^{-\alpha_2-1} + \mathcal{O}\left(s^{-\beta_2-1}\right)\right),\\ c &= \left(A_2 \frac{(\alpha_2+1)\alpha_2}{2}(s-q_{\alpha_2})^{-\alpha_2-2} + \mathcal{O}\left((s-q_{\beta_2})^{-\beta_2-2}\right)\right)\end{aligned}$$

Further evaluation gives for s large enough:

$$\int_{\frac{-s}{2}}^{\frac{s}{2}} f_1(x)dx = 1 - A_1\left(\frac{s}{2}\right)^{-\alpha_1} - A_1^L\left(\frac{s}{2}\right)^{-\alpha_1^L} + \mathcal{O}\left(s^{-\beta_1}\right) + \mathcal{O}\left(s^{-\beta_1^L}\right),$$

where $\beta > \max\{\alpha_1, \alpha_1^L\}$.

$$\int_{\frac{-s}{2}}^{\frac{s}{2}} x f_1(x)dx = \int_{\frac{-s}{2}}^{-\sqrt{s}} x f_1(x)dx + \int_{-\sqrt{s}}^{0} x f_1(x)dx + \int_{0}^{\sqrt{s}} x f_1(x)dx + \int_{\sqrt{s}}^{\frac{s}{2}} x f_1(x)dx = (1) + (2) + (3) + (4),$$

and

$$\int_{\frac{-s}{2}}^{\frac{s}{2}} x^2 f_1(x)dx = \int_{\frac{-s}{2}}^{-\sqrt[3]{s}} x^2 f_1(x)dx + \int_{-\sqrt[3]{s}}^{\sqrt[3]{s}} x^2 f_1(x)dx + \int_{\sqrt[3]{s}}^{\frac{s}{2}} x^2 f_1(x)dx = (5) + (6) + (7).$$

Substitution of $y = x^2$ for the integral expression in (3) gives the following:

$$(3) = \int_0^{\sqrt{s}} x f_1(x)dx = \int_0^s \frac{1}{2} f_1\left(\sqrt{y}\right)dy \le \frac{s}{2}C_1 = \mathcal{O}(s),$$

where we use that $f_1 \le C_1$ a.s. In the same way we can show that expression (2) $= \mathcal{O}(s)$. (4) can be calculated

as well and we obtain:

$$\begin{aligned}(4) &= \int_{\sqrt{s}}^{\frac{s}{2}} x \frac{d\mathbb{P}(X_1 \le x)}{dx} dx \\ &= \int_{\sqrt{s}}^{\frac{s}{2}} x \frac{d}{dx}\left(A_1 x^{-\alpha_1} + \mathcal{O}\left(x^{-\beta_1}\right)\right) dx \\ &= \int_{\sqrt{s}}^{\frac{s}{2}} \left(-\alpha_1 A x^{-\alpha_1} + \mathcal{O}\left(x^{-\beta_1}\right)\right) dx \\ &\overset{*}{=} \left[\frac{-\alpha_1}{-\alpha_1+1} A_1 x^{-\alpha_1+1} + \mathcal{O}\left(x^{-\beta_1+1}\right)\right]_{\sqrt{s}}^{\frac{s}{2}} \\ &= \mathcal{O}\left(s^k\right),\end{aligned}$$

where $k = \max\left\{-\alpha_1+1, -\frac{\alpha_1}{2}+\frac{1}{2}\right\} < 1$. Note that $*$ only holds for $\alpha_1, \beta_1 \neq 1$. If $\alpha_1 = 1$ or $\beta_1 = 1$ we would get a $\mathcal{O}(\log s) < \mathcal{O}(s)$ and this will therefore not change the end result. Using a similar derivation we obtain that expression $(1) = \mathcal{O}\left(s^l\right)$, where $l = \max\left\{-\alpha_1^L+1, -\frac{\alpha_1^L}{2}+\frac{1}{2}\right\} < 1$.

Substitution of $y = x^3$ for the integral expression in (6) gives us the following:

$$(6) = \int_{-s}^{s} \frac{1}{3} f_1(\sqrt[3]{y}) dy \le \frac{2s}{3} C_1 = \mathcal{O}(s),$$

where we use that $f_1 \le C_1$ a.s. again. (5) and (7) can be approximated for large s and for (7) this gives us:

$$\begin{aligned}(7) &= \int_{\sqrt[3]{s}}^{\frac{s}{2}} x^2 \frac{d\mathbb{P}(X_1 \le x)}{dx} dx \\ &= \int_{\sqrt[3]{s}}^{\frac{s}{2}} x^2 \frac{d}{dx}\left(\left(1 - A_1 x^{-\alpha_1}\right) + \mathcal{O}\left(x^{-\beta_1}\right)\right) dx \\ &= \int_{\sqrt[3]{s}}^{\frac{s}{2}} \left(\alpha_1 A_1 x^{-\alpha_1+1} + \mathcal{O}\left(x^{-\beta_1+1}\right)\right) dx \\ &\overset{*}{=} \left[\frac{-\alpha_1}{-\alpha_1+2} A_1 x^{-\alpha_1+2} + \mathcal{O}\left(x^{-\beta_1+2}\right)\right]_{\sqrt[3]{s}}^{\frac{s}{2}} \\ &= \frac{\alpha_1}{\alpha_1-2} A_1 \left(\left(\frac{s}{2}\right)^{-\alpha_1+2} - s^{-\frac{\alpha_1}{3}+\frac{2}{3}} + \mathcal{O}\left(s^{-\beta_1+2}\right) + \mathcal{O}\left(s^{-\frac{\beta_1}{3}+\frac{2}{3}}\right)\right) \\ &= \mathcal{O}\left(s^m\right),\end{aligned}$$

where $m = \max\left\{-\alpha_1+2, -\frac{\alpha_1}{3}+\frac{2}{3}\right\} < 2$. Note that $*$ only holds for $\alpha_1, \beta_1 \neq 2$. $\alpha_1 = 2$ or $\beta_1 = 2$ would give terms of order $\mathcal{O}(\log s) < \mathcal{O}\left(s^2\right)$. The derivation of (5) is similar and $(5) = \mathcal{O}\left(s^n\right)$ where $n = \max\left\{-\alpha_1^L+2, -\frac{\alpha_1^L}{3}+\frac{2}{3}\right\} < 2$. Combining these results and using the expressions for a, b and c gives for I_A:

$$\begin{aligned}\mathrm{I}_A &= \left(1 - A_2(s-x)^{-\alpha_2} + \mathcal{O}\left((s-x)^{-\beta_2}\right)\right) \cdot \left(1 - A_1\left(\frac{s}{2}\right)^{-\alpha_1} - A_1^L\left(\frac{s}{2}\right)^{-\alpha_1^L}\right) - \left(A_2\alpha_2 s^{-\alpha_2-1} + \mathcal{O}\left(s^{-\beta_2-1}\right)\right) \cdot \mathcal{O}(s) \\ &\quad - \left(A_2 \frac{(\alpha_2+1)\alpha_2}{2}(s-q_{\alpha_2})^{-\alpha_2-2} + \mathcal{O}\left((s-q_{\beta_2})^{-\beta_2-2}\right)\right) \cdot (\mathcal{O}(s) + \mathcal{O}\left(s^m\right) + \mathcal{O}\left(s^n\right)) \\ &= 1 - A\left(\frac{s}{2}\right)^{-\alpha} + \mathcal{O}\left(s^{-\beta}\right),\end{aligned}$$

where $\beta > \alpha = \min\{\alpha_1, \alpha_1^L\}$, since $m, n \le 2$.

II_A is easier to compute and gives:

$$\begin{aligned}\mathrm{II}_A &= F_2\left(\frac{s}{2}\right) \int_{\frac{-s}{2}}^{\frac{s}{2}} f_1(x) dx \\ &= \left(1 - A_2\left(\frac{s}{2}\right)^{-\alpha_2} + \mathcal{O}\left(s^{-\beta}\right)\right) \cdot \left(1 - A_1\left(\frac{s}{2}\right)^{-\alpha_1} - A_1^L\left(\frac{s}{2}\right)^{-\alpha_1^L} + \mathcal{O}\left(s^{-\beta}\right)\right) \\ &= 1 - A\left(\frac{s}{2}\right)^{-\alpha} + \mathcal{O}\left(s^{-\beta}\right),\end{aligned}$$

where $\beta > \alpha = \min\{\alpha_1, \alpha_1^L\}$. Therefore we have that:

$$\mathbb{P}(A) = \mathrm{I}_A - \mathrm{II}_A = \mathcal{O}\left(s^{-\beta}\right),$$

with $\beta > \alpha$. $\mathbb{P}(B)$ can be calculated in the same way. This also gives us:

$$\mathbb{P}(B) = \mathrm{I}_B - \mathrm{II}_B = \mathcal{O}\left(s^{-\beta}\right).$$

$\mathbb{P}(D)$ and $\mathbb{P}(E)$ are of lower order. Namely:

$$\mathbb{P}(D) \leq \mathbb{P}\left(X_1 \leq \frac{-s}{2}, X_2 > \frac{s}{2}\right) = \mathcal{O}\left(s^{-\alpha_1^L}\right)\mathcal{O}\left(s^{-\alpha_2}\right) = \mathcal{O}\left(s^{-\alpha_1^L-\alpha_2}\right),$$

and similarly $\mathbb{P}(E) = \mathcal{O}\left(s^{-\alpha_1-\alpha_2^L}\right)$. Adding up all these probabilities gives us:

$$\mathbb{P}(X_1 + X_2 > s) = 1 - \mathbb{P}(A) - \mathbb{P}(B) - \mathbb{P}(C) - \mathbb{P}(D) - \mathbb{P}(E) = A_1 2^{\alpha_1} s^{-\alpha_1} + \mathcal{O}\left(s^{-\beta}\right),$$

where $\beta > \alpha_1$. Therefore, the largest upper tail index survives in the upper tail. □

Note that also in the lower tail the largest lower tail index survives, the derivation is similar and we get:

$$\mathbb{P}(X_1 + X_2 \leq -s) = A_1^L 2^{\alpha_1^L} s^{-\alpha_1^L} + \mathcal{O}\left(s^{-\beta}\right),$$

where $\beta > \alpha_1^L$.
If we drop the assumption that one of the tail indices is larger than the other we get $\alpha_1 = \alpha_2 > 0$ and the result will stay the same. The proof is similar to the proof of the $\gamma_1 \neq \gamma_2$ case (i.e. $\alpha_1 \neq \alpha_2$). We can combine this result with the result of section 3.4.1 to obtain the following corollary.

Corollary 3.9. *If $X = \sum_{i=1}^n c_i X_i$, where $c_i \neq 0$, X_i has lower tail index γ_i^L and upper tail index γ_i^U, then X has tail index γ, where $\gamma = \max_{1\leq i\leq n}\{\gamma_i\}$, with:*

$$\gamma_i = \begin{cases} \gamma_i^L, & \text{if } c_i < 0, \\ \gamma_i^U, & \text{if } c_i > 0. \end{cases}$$

3.4.3. Infinite sum of heavy-tailed random variables

Theorem 3.10. *Suppose that we have a countable infinitely weighted sum of independent and identically distributed random variables $y = \sum_{j=0}^{\infty} c_j x_j$, where x_j follows a heavy-tailed distribution with finite expectation (i.e. lower tail index $\gamma^L < 1$, upper tail index $\gamma^U < 1$), almost surely continuous density function and $\sum_{j=0}^{\infty} c_j = C < \infty$.*
Then y follows a heavy-tailed distribution with upper tail index γ, where

$$\gamma = \begin{cases} \gamma^L, & \text{if } c_j < 0 \ \forall j, \\ \gamma^U, & \text{if } c_j > 0 \ \forall j, \\ \max\{\gamma^L, \gamma^U\}, & \text{else.} \end{cases} \tag{3.12}$$

Proof.

For some s large enough and $\beta > \frac{1}{\gamma}$ we have that:

$$\begin{aligned}
\mathbb{P}\left(\sum_{j=0}^{\infty} c_j x_j > s\right) &= \mathbb{P}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right) \\
&= \mathbb{E}\left[\mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right)\right] \\
&\overset{(1)}{=} \mathbb{E}\left[\lim_{n\to\infty} \mathbb{1}\left(\sum_{j=0}^{n} c_j x_j > s\right)\right] \\
&\overset{(2)}{=} \lim_{n\to\infty} \mathbb{E}\left[\mathbb{1}\left(\sum_{j=0}^{n} c_j x_j > s\right)\right] \\
&= \lim_{n\to\infty} \mathbb{P}\left(\sum_{j=0}^{n} c_j x_j > s\right) \\
&\overset{(3)}{\approx} \lim_{n\to\infty} \left(A_n s^{-\frac{1}{\gamma}} + \mathcal{O}\left(s^{-\beta}\right)\right) \\
&= A s^{-\frac{1}{\gamma}} + \mathcal{O}\left(s^{-\beta}\right),
\end{aligned}$$

for some $A \in \mathbb{R}$. $A = \lim_{n\to\infty} A_n$ exists, since the probability is bounded by 1. Here (3) follows directly from Corollary 3.9. For (1) we will show that:

$$\lim_{n\to\infty} \mathbb{1}\left(\sum_{j=0}^{n} c_j x_j > s\right) = \mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right) \text{ a.s.}$$

For (2) we will use dominated convergence, but first we will show (1).

To show (1) we will distinguish the following two cases:

(a) $\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j = L > s$ (then $\mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right) = 1$);

(b) $\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j = L < s$ (then $\mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right) = 0$).

From the definition of the limit we have that for $\epsilon > 0$ arbitrary we can find an $N \in \mathbb{N}$ such that $\forall n \geq N$: $L - \epsilon < \sum_{j=0}^{n} c_j x_j < L + \epsilon$.
First assume (a).
We choose $\epsilon = \frac{1}{2}|L - s|$ and pick $N \in \mathbb{N}$ such that $\forall n \geq N$: $\sum_{j=0}^{n} c_j x_j > L - \epsilon > s$.
Therefore

$$\lim_{n\to\infty} \mathbb{1}\left(\sum_{j=0}^{n} c_j x_j > s\right) = 1 = \mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right).$$

Similarly when (b) holds we obtain that:

$$\lim_{n\to\infty} \mathbb{1}\left(\sum_{j=0}^{n} c_j x_j > s\right) = 0 = \mathbb{1}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j > s\right).$$

The equality therefore holds under (a) and under (b). We will show that (a) or (b) happens almost surely by showing that the limit L exists almost surely. Note that the case $L = s$ happens with probability 0, since the density function is bounded a.s. That limit L exists almost surely can be shown by showing that:

$$\mathbb{E}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j\right) < \infty.$$

We know that:

$$\mathbb{E}\left(\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j\right) \leq \mathbb{E}\left(\left|\lim_{n\to\infty} \sum_{j=0}^{n} c_j x_j\right|\right) \leq \mathbb{E}\left(\lim_{n\to\infty} \sum_{j=0}^{n} |c_j| \cdot |x_j|\right) \overset{*}{=} \lim_{n\to\infty} \mathbb{E}\left(\sum_{j=0}^{n} |c_j| \cdot |x_j|\right) < \infty,$$

where (*) is possible because $\mathbb{E}\left[\sum_{j=0}^{n}|c_j|\cdot|x_j|\right]$ is uniformly bounded by:

$$\mathbb{E}\left[\sum_{j=0}^{n}|c_j|\cdot|x_j|\right]=\sum_{j=0}^{n}|c_j|\mathbb{E}\left[|x_j|\right]\leq\sum_{j=0}^{n}|c_j|\mathbb{E}\left[|x_1|\right]=\mathbb{E}\left[|x_1|\right]C<\infty.$$

Since (a) or (b) holds almost surely, we have that:

$$\lim_{n\to\infty}\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)=\mathbb{1}\left(\lim_{n\to\infty}\sum_{j=0}^{n}c_jx_j>s\right)\text{ a.s.}$$

Now we will show (2):
Define $g=\mathbb{1}\left(\sum_{j=0}^{\infty}|c_jx_j|>s\right)$, then clearly $|\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)|\leq g$ and g is measurable. Also $\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)$ converges to $\lim_{n\to\infty}\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)$ pointwise and from the dominated convergence theorem:

$$\lim_{n\to\infty}\mathbb{E}\left[\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)\right]=\mathbb{E}\left[\lim_{n\to\infty}\mathbb{1}\left(\sum_{j=0}^{n}c_jx_j>s\right)\right]$$

So, now we have that $\mathbb{P}\left(\sum_{j=0}^{\infty}c_jx_j>s\right)=As^{-\frac{1}{\gamma}}+\mathcal{O}\left(s^{-\beta}\right)$ for s large enough and $\beta>\frac{1}{\gamma}$. Thus, using Proposition 2.1, we can conclude that $y=\sum_{j=0}^{\infty}c_jx_j$ has tail index γ. □

Note that the convergence towards the tail can be significantly slower than the convergence in the original errors. This may make it more difficult to estimate the tail parameter γ on this sum than on the errors, especially for small-sized data sets. In the next section we will extend this result to VAR time series with heavy-tailed input errors.

3.5. Tail index of VAR time series

The following theorem relates the time series to the input errors of a VAR(p) model. We will show that it follows directly from the result of Theorems 3.3 and 3.10.

Theorem 3.11. *Suppose that $y^{(t)}\in\mathbb{R}^n$ is described by a stationary VAR(p) model through:*

$$y^{(t)}=A_1y^{(t-1)}+A_2y^{(t-2)}+\cdots+A_py^{(t-p)}+\epsilon^{(t)},$$

where the errors are independent and identically distributed with respect to time. Then component $y_i^{(t)}$ can be written as a countable infinite sum of the errors, namely:

$$y_i^{(t)}=\sum_{k=1}^{n}\sum_{p=0}^{\infty}c_{p,k}\epsilon_k^{(t-p)},\tag{3.13}$$

which has upper tail index $\gamma_i=\max_{k\leq n}\left\{\max\left\{\mathbb{1}_{i,k}^U\gamma_k^U,\mathbb{1}_{i,k}^L\gamma_k^L\right\}\right\}$, where

$$\mathbb{1}_{i,k}^U=\begin{cases}0, & \text{if } c_{p,k}\leq 0\ \forall p,\\ 1, & \text{else,}\end{cases}\tag{3.14}$$

and

$$\mathbb{1}_{i,k}^L=\begin{cases}0, & \text{if } c_{p,k}\geq 0\ \forall p,\\ 1, & \text{else.}\end{cases}\tag{3.15}$$

(i.e. the tail index of y is equal to the largest of the upper and lower tail indices of the tails that influence y.)

Proof. First use Theorem 3.3, which states that any stationary n-dimensional VAR(p) model can be written as a VMA(∞) model. From this theorem it follows that every component of this VAR(p) time series $y_i^{(t)}$ can be written as a countably infinite sum of error components and therefore equation (3.13) holds.

$$y_i^{(t)}=\sum_{k=1}^{n}\sum_{p=0}^{\infty}c_{p,k}\epsilon_k^{(t-p)},\tag{3.16}$$

where for every k the errors $\epsilon_k^{(t)}$ are independent and identically distributed with respect to t. From Theorem 3.10 we have that $\sum_{p=0}^{\infty} c_{p,k}\epsilon_k^{(t-p)}$ has tail index γ_k, where

$$\gamma_k = \begin{cases} 0, & \text{if } c_{p,k} = 0 \ \forall p, \\ \gamma_k^L, & \text{if } c_{p,k} \leq 0 \ \forall p \text{ and } \exists p \text{ such that } c_{p,k} \neq 0, \\ \gamma_k^U, & \text{if } c_{p,k} \geq 0 \ \forall p \text{ and } \exists p \text{ such that } c_{p,k} \neq 0, \\ \max\{\gamma_k^L, \gamma_k^U\}, & \text{else.} \end{cases} \tag{3.17}$$

From Corollary 3.9 we have that the tail index of $y_i^{(t)}$ equals $\gamma = \max_{k\leq n}\{\gamma_k\}$. Furthermore $\gamma_k = \max\left\{\mathbb{1}_{i,k}^U\gamma_k^U, \mathbb{1}_{i,k}^L\gamma_k^L\right\}$ and it follows that the tail index is given by:

$$\gamma = \max_{k\leq n}\{\gamma_k\} = \max_{k\leq n}\left\{\max\left\{\mathbb{1}_{i,k}^U\gamma_k^U, \mathbb{1}_{i,k}^L\gamma_k^L\right\}\right\}.$$

□

Example 3.2. *Suppose that factors* $f^{(t)} = \begin{pmatrix} f_1^{(t)} & f_2^{(t)} \end{pmatrix}^T$ *are described by a VAR*(1) *model through:*

$$f^{(t)} = \begin{pmatrix} -0.5 & -0.4 \\ -0.4 & -0.5 \end{pmatrix} f^{(t-1)} + \epsilon^{(t)},$$

with $\epsilon_i^{(t)}$ *following a heavy-tailed distribution with upper and lower tail index* $\gamma_1^U = \gamma_1^L = 0.1$ *and* $\gamma_2^U = \gamma_2^L = 0.5$. *Then we can diagonalise A as follows* $A = PDP^{-1}$, *where:*

$$P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix},$$

and

$$D = \begin{pmatrix} -0.1 & 0 \\ 0 & -0.9 \end{pmatrix}.$$

Using Theorem 3.3 we can write:

$$f^{(t)} = \sum_{p=0}^{\infty} A^p\epsilon^{(t-p)} = \sum_{p=0}^{\infty} A^p\epsilon^{(t-p)} = \sum_{p=0}^{\infty} PD^pP^{-1}\epsilon^{(t-p)}.$$

After working out the matrix products, we can write:

$$f_1^{(t)} = 0.5\sum_{p=0}^{\infty}((-0.1)^p + (-0.9)^p)\epsilon_1^{(t-p)} + 0.5\sum_{p=0}^{\infty}(-(-0.1)^p + (-0.9)^p)\epsilon_2^{(t-p)},$$

and:

$$f_2^{(t)} = 0.5\sum_{p=0}^{\infty}(-(-0.1)^p + (-0.9)^p)\epsilon_1^{(t-p)} + 0.5\sum_{p=0}^{\infty}((-0.1)^p + (-0.9)^p)\epsilon_2^{(t-p)}.$$

In all the sums the coefficients take positive as well as negative values and depend on ϵ_1 *and* ϵ_2. *Hence, the upper and lower tail indices of both factors are given by* $\gamma^U = \gamma^L = \max\{\gamma_1^U, \gamma_1^L, \gamma_2^U, \gamma_2^L\} = 0.5$.

In Example 3.2 we see that the upper and lower tail index of every factor can depend on every individual upper and lower tail index of the errors. When the matrix A contains elements that are equal to 0, this result can be very different, which is illustrated by Example 3.3.

Example 3.3. *Suppose that factors* $f^{(t)} = \begin{pmatrix} f_1^{(t)} & f_2^{(t)} \end{pmatrix}^T$ *are described by a VAR*(1) *model through:*

$$f^{(t)} = \begin{pmatrix} 0.5 & 0 \\ 0 & -0.5 \end{pmatrix} f^{(t-1)} + \epsilon^{(t)},$$

with $\epsilon_1^{(t)}$ *and* $\epsilon_2^{(t)}$ *following a heavy-tailed distribution with upper and lower tail index* $\gamma_1^U = \gamma_1^L = 0.1$ *and* $\gamma_2^U = \gamma_2^L = 0.5$, *respectively.*
For factor 1 *we then have that:*

$$f_1^{(t)} = \sum_{p=0}^{\infty} 0.5^p\epsilon_1^{(t-p)},$$

and factor 1 *follows a heavy-tailed distribution with* $\gamma^U = \gamma_1^U = 0.1$ *and* $\gamma^L = \gamma_1^L = 0.1$, *since* ϵ_2 *does not influence* f_1 *and* $0.5^p > 0 \; \forall p$. *Similarly factor* 2 *is given by:*

$$f_2^{(t)} = \sum_{p=0}^{\infty} (-0.5)^p \epsilon_2^{(t-p)},$$

and factor 2 *follows a heavy-tailed distribution with* $\gamma^U = \gamma^L = \max\{\gamma_2^U, \gamma_2^L\} = 0.5$, *since* ϵ_1 *does not influence* f_2 *and* $(-0.5)^p$ *takes as well positive as negative values.*

Due to the fact that the matrix can be decoupled in Example 3.3, the tail index of factor i only depends on the tail index of ϵ_i. When we would not have had zero entries, this would not have been the case, since the indicators in equations (3.14) (3.15) would be zero.

3.6. Estimation of tail indices of VAR models

In this section we will estimate the tail index on the errors of a VAR model and directly on the factors that are described by the VAR model. If indeed the tail index of time series modelled as a VAR model is the same as the tail index of its errors, we should see convergence towards this tail index in the estimation when the sample size increases. First we will estimate the tail index of a simple AR(1) time series and in the second part of this section we will do this for a VAR(1) time series.

3.6.1. Tail index of AR time series

Let us consider a simple AR(1) model with $\phi = -0.9, -0.6, \cdots, 0.9$, recall Definition 3.1. Here, the errors e_i are simulated from a double Pareto distribution with upper and lower tail index equal and ranging from 0.1 to 0.5 again. The density function of this distribution is the average of a Pareto and a mirrored Pareto density such that this distribution is symmetric around 0 if the upper and lower tail are equal. Therefore, we also have that $\mathbb{E}[e_i] = 0 \; \forall i$. Note that $\phi = 0$ corresponds to the time series variable being equal to the errors and corresponds to direct estimation of γ on the errors.

The estimation of the upper tail indices of 500 simulated AR(1) time series with 600 data points is given in Figure 3.1. In the estimation we only consider the positive data $\{X_1, \cdots, X_n\}$, since the estimators are based on having only positive (or only negative) data. We will have $\mathbb{E}[n] = 300$ and therefore use the estimator we found in section 2.5.6 based on sample size 300. Similarly, we simulated the time series for sample sizes 2000 and 20000 of which the results are given in Figures 3.2 and 3.3. Here we took the average over 200 simulations, since the estimator variance is smaller for larger sample sizes. In Tables C.4, C.5 and C.6 in the appendix we show the MSEs of these estimates.

Figure 3.1: **Estimated tail index of AR time series**

This figure shows the estimates of the tail index of the factor described by an AR(1) model as a function of the tail index of the error, which ranges from 0.1 to 0.5. The value of the AR coefficient ϕ ranges from -0.9 to 0.9 and the sample size of the positive data is on average approximately 300.

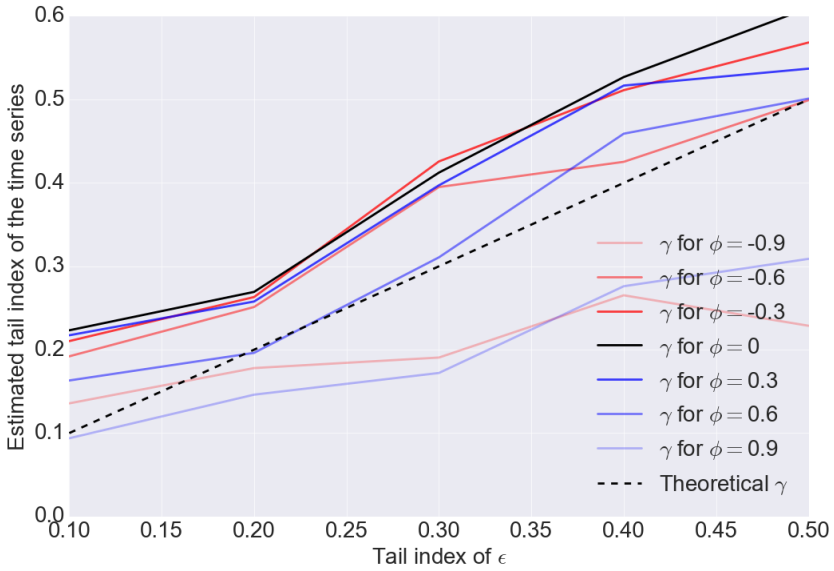


Figure 3.2: **Estimated tail index of AR time series**

This figure shows the estimates of the tail index of the factor described by an AR(1) model as a function of the tail index of the error, which ranges from 0.1 to 0.5. The value of the AR coefficient ϕ ranges from -0.9 to 0.9 and the sample size of the positive data is on average approximately 1000.

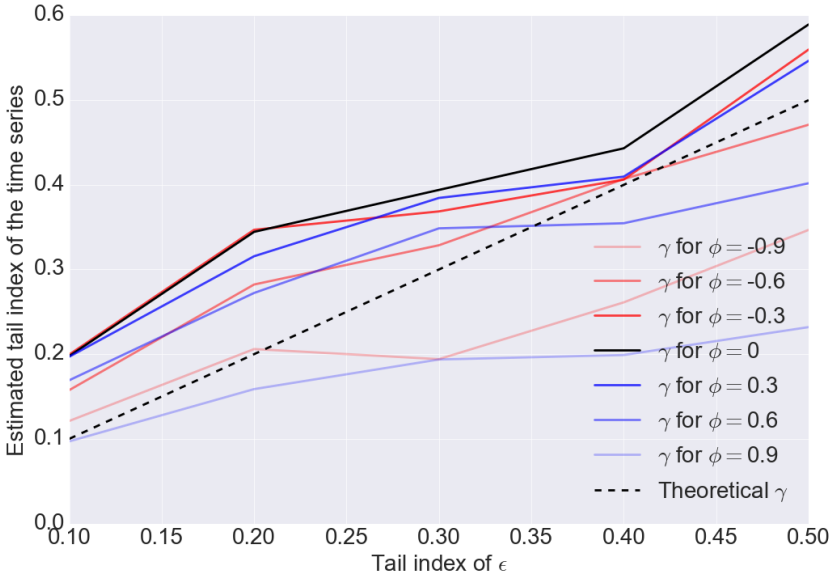
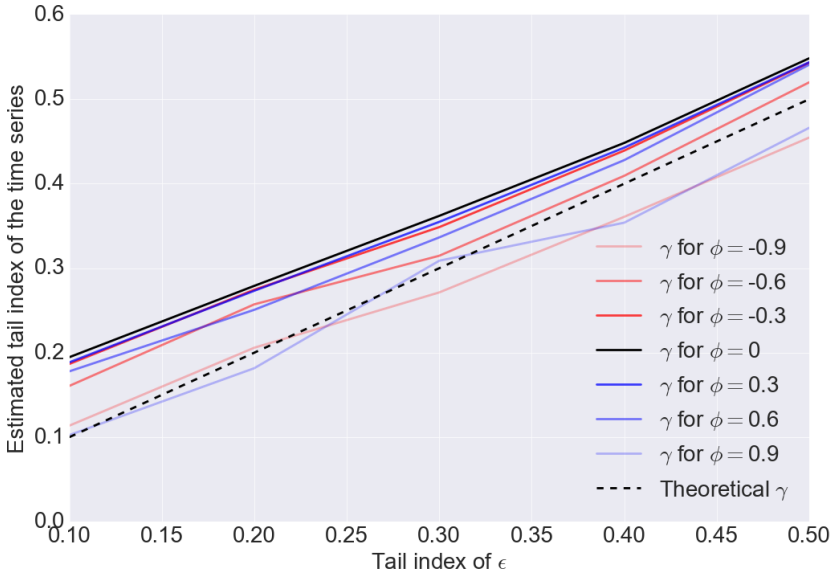


Figure 3.3: **Estimated tail index of AR time series**

This figure shows the estimates of the tail index of the factor described by an AR(1) model as a function of the tail index of the error, which ranges from 0.1 to 0.5. The value of the AR coefficient ϕ ranges from -0.9 to 0.9 and the sample size of the positive data is on average approximately 10000.



In Figures 3.1, 3.2 and 3.3 we see that the estimate of the tail index decreases when ϕ moves away from zero. This effect is stronger when the sample size is smaller and the estimates are closer to the theoretical value for larger sample sizes. In Tables C.4, C.5 and C.6 in the appendix we can also see that the MSE decreases for larger sample sizes. This is in line with the fact that the theoretical tail index of the AR model equals the tail index of the error.

In the tables of Appendix B, the appendix with results of the previous chapter, we can see that the tail index estimator has a positive bias for the Pareto distribution. Therefore, we expect to have a positive bias for $\phi = 0$ as well, since the time series is then equal to the error series, which follows a double Pareto distribution. For the $|\phi| \geq 1$ case the distribution of the time series variable is not stationary, meaning that we cannot speak of a tail index. For $|\phi|$ between 0 and 1 the time series is equal to a weighted sum of the errors. Extreme observations for individual errors will therefore be less prominent in the resulting time series. For finite samples, this may lower the tail index estimate even though the theoretical value is unchanged. Since the $\phi = 0$ case gives a positive bias, at first we expect the bias to decrease as $|\phi|$ increases. When $|\phi|$ increases too much, however, the bias will be negative and increase in absolute value. The performance of the estimator will then deteriorate.

This behaviour can be seen in all three figures. As $|\phi|$ increases slightly, see the $\phi = \pm 0.3$ lines, the bias generally gets smaller. The $\phi = \pm 0.3$ lines are mostly below the $\phi = 0$ lines. From Tables C.4, C.5 and C.6 we also see that the MSE decreases. This bias decreases even more when $|\phi|$ increases by a larger amount. The best estimates, in MSE, are the estimates for the $|\phi| = 0.6$ case. This result holds for double Pareto distributed errors, but not for arbitrary distributions. We should also note here that the estimators are designed for independent samples. When $\phi \neq 0$ we however have autocorrelation and therefore the observations are not mutually independent, which might affect the performance of the estimators.

3.6.2. Tail index of VAR time series

In Examples 3.2 and 3.3 we discussed the differences between the final tail indices for different VAR models. In Figures 3.4 and 3.5 the results are given for the resulting upper tail indices when we plug in double Pareto errors with equal upper and lower tail indices $\gamma_1 = 0.1, \cdots, 0.5$ and $\gamma_2 = 0.5$ for $\epsilon_1^{(t)}$ and $\epsilon_2^{(t)}$.

Figure 3.4: **Estimated tail indices in Example 3.2**

This figure shows the estimates of the tail indices of the factors described by the VAR model in Example 3.2. Here, the tail index of the first error varies from 0.1 to 0.5 and of the second error is 0.5.

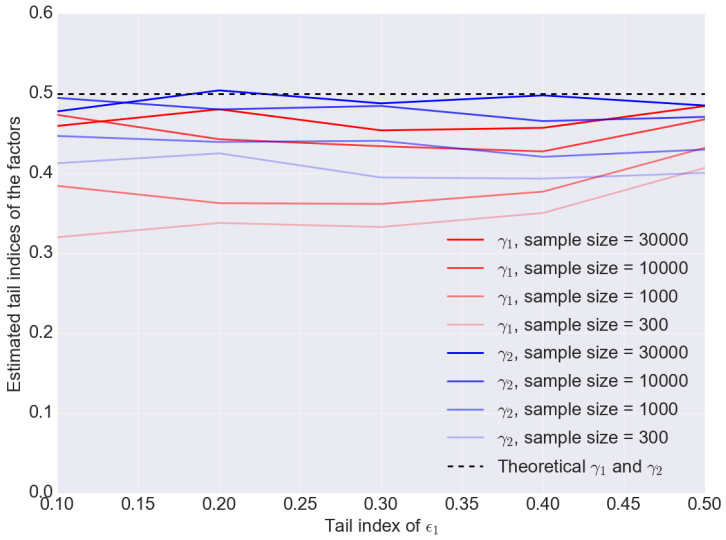
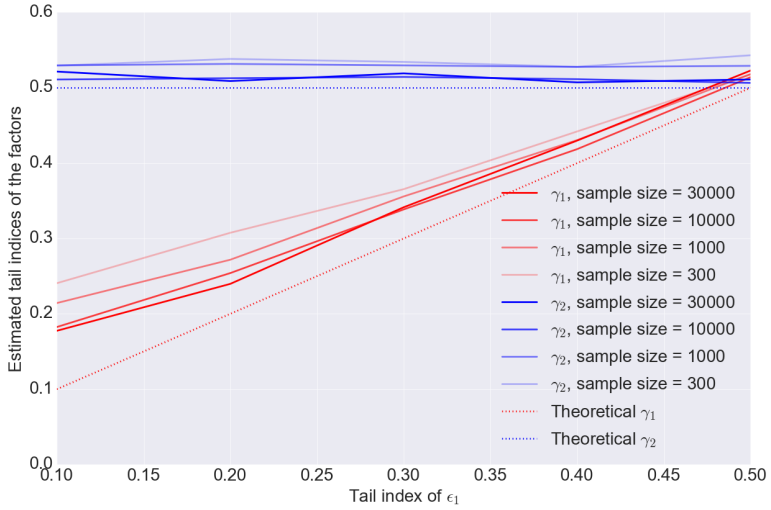


Figure 3.5: **Estimated tail indices in Example 3.3**

This figure shows the estimates of the tail indices of the factors described by the VAR model in Example 3.3. Here, the tail index of the first error varies from 0.1 to 0.5 and of the second error is 0.5.



The estimation on time series of 600 points can have a large bias see for example the line for γ_1 in Figure 3.4. In both figures we see that when the sample size increases, that the average tail index gets closer to the theoretical values, which is what we expect.

3.7. Conclusion

At Ortec Finance, a dynamic factor model is used to describe the behaviour of economic and financial variables. The model makes use of statistical factors that are described by a VAR model. We found that, under

some assumption, a countable sum of heavy-tailed random variables has the same tail index as the largest tail index of these random variables. As a result, we obtain that in VAR models with heavy-tailed errors the tail index of a factor equals the largest tail index of the errors, which influences this factor. Therefore, if we model the input errors as heavy-tailed random variables and if all factors influence each other through the matrix product, the factors will all have the same tail index. Since all other financial variables depend on the factors through a regression or depend on it indirectly we will model all financial variables with the same tail index. We have no indication that all financial variables have the same tail index and therefore we should be careful in using this approach for modelling the tails.

Simulations support this theoretical result, but the convergence of estimations towards this theoretical tail index can be slower for the factors than for the errors.

4

Modelling tail dependence using copulas

4.1. Introduction

In financial data stocks, bonds, and many other financial variables often show high correlation during crisis times, while in normal times this is not necessarily the case, see Aloui et al. (2011). This indicates that we should not only use one constant correlation coefficient to model the dependence between financial variables. Using a single, constant correlation index can underestimate the dependence in the tails of distributions and therefore underestimates the corresponding risks. By the use of copulas, a more advanced way to capture dependence relations, we aim to model the dependence of different random variables on the whole range of possible outcomes of this random variable. A copula is used to link the marginal distributions of several variables to a multivariate distribution and this is given in equation (4.1), see Nelsen (1999).

$$F(x_1, x_2, \cdots, x_n) = C(F_{x_1}(x_1), F_{x_2}(x_2), \cdots, F_{x_n}(x_n)), \tag{4.1}$$

where F is the multivariate distribution function and F_{x_i} is the marginal distribution function of variable x_i. As we can see here, this decouples the dependence and the individual behaviour of the x_i's. The copula captures the multivariate behaviour of $x_1, \cdots, x_n$ and the marginal distribution functions describe the individual behaviour of each x_i. In this thesis we use static copulas that are suitable when the multivariate distributions are static, meaning that this distribution itself is constant in time. This assumption does not generally hold, but in many applications it can be an accurate approximation. In the modelling of Ortec Finance the VAR model captures the dynamical part of the factors and all underlying errors that we will analyse are assumed to be independent of time.

In this chapter we will first introduce copulas and some of their properties. We will also elaborate on some dependence measures, especially tail dependence. Subsequently parametric copulas, such as Archimedean copulas, are studied and we will show this copula approach has some possible limitations. We proceed with a semiparametric approach of which it is a challenge to include tail dependence. This problem will be solved for the 2-dimensional setting and we will find ways to estimate the tail dependence coefficients. Lastly, we will use this copula with this estimated tail dependence to generate tail dependent random variables and ultimately tail dependent factors.

4.2. Copulas and properties

A copula has to satisfy some conditions that are formally defined by Nelsen (1999) and its definition is given below.

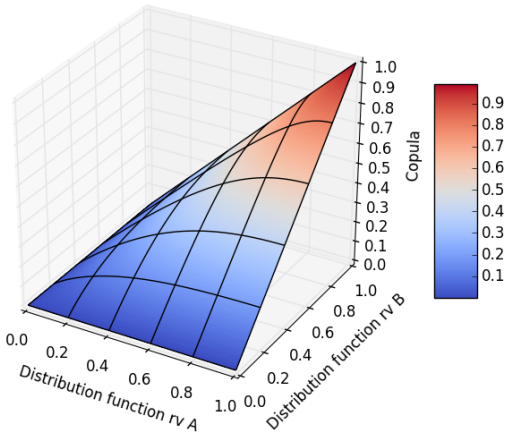
Definition 4.1 (Copula)**.** *$C : [0,1]^n \to [0,1]$ is called an n-dimensional copula if and only if the following holds:*
(1) $\forall u = (u_1, u_2, \cdots, u_n)^T \in [0,1]^n, C(u_1, \cdots, u_n) = 0$ *if* $\exists 1 \leq i \leq n$ *such that* $u_i = 0$ *(i.e. C is grounded).*
(2) C has marginals C_i satisfying $\forall i = 1, \cdots, n, \forall u \in [0,1] : C_i(u) = u$*, (here* $C_i(u) = C(1, \cdots, 1, u, 1, \cdots, 1)$*, where u is on position i).*
(3) $\forall u = (u_1, u_2, \cdots, u_n)^T$, $v = (v_1, v_2, \cdots, v_n)^T \in [0,1]^n$*, such that* $\forall j \in \{1, \cdots, n\}$*:* $v_j \leq u_j$,

$$V_C(B) := D_{u_n}^{v_n} \cdots D_{u_1}^{v_1} C(t_1, \cdots t_n) \geq 0,$$

where $B = ([u_1, v_1] \times \cdots \times [u_n, v_n])$, $t \in [0,1]^n$ *and*
$D_{u_i}^{v_i} C(t_1, \cdots t_n) = C(t_1, \cdots, t_{j-1}, v_j, t_{j+1}, \cdots, t_n) - C(t_1, \cdots, t_{j-1}, u_j, t_{j+1}, \cdots, t_n)$ *(i.e. C is n-increasing).*

Figure 4.1: **Copula example.**

This figure shows an example of a widely used copula, namely the Gumbel copula with model parameter $\theta = 2$.



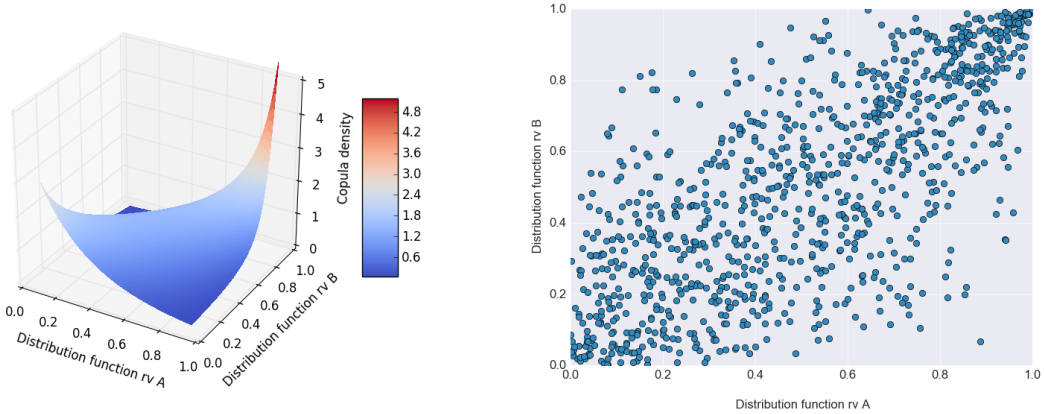
In Figure 4.1 we see an example of a 2-dimensional copula. This is a Gumbel copula, which will be further discussed in section 4.5. It is difficult to distinguish between different copulas, since all copulas are identical on the boundary of their domain. It is easier to distinguish between different copulas using the copula density, which does not necessarily exist everywhere. For points where the copula is continuously differentiable it is given by:

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}. \tag{4.2}$$

The copula density corresponding to the copula in Figure 4.1 is given in Figure 4.2. In this figure we also show a sample drawn from this copula. This sample and density give more insight in the behaviour of the underlying distributions. The copula density tends to infinity for $u_1, u_2 \to 1$ and we therefore see some clustering close to $(1,1)$ in the sample. This indicates that there is some extra dependence in the upper tail of the distribution, which suggests the possible phenomenon of tail dependence. We will introduce the definition of tail dependence and elaborate on this topic in section 4.3.

Figure 4.2: **Copula density example.**

The left figure shows an example of the density of a widely used copula, namely the Gumbel copula with model parameter $\theta = 2$. In the right figure we give a sample of 1000 independent pairs of random variables from the same copula.



(a) Copula density

(b) Sample from copula

Computations show that its density function is not defined for the points $(0,0)$ and $(1,1)$, since the density diverges to infinity there. This divergence in $(0,0)$, $(0,1)$, $(1,0)$ and/or $(1,1)$ is common for many parametric copulas.

A copula is a function, which is used to link the marginal distributions to the multivariate distribution. Whether this link is unique is discussed in Sklar (1973) and they proved that this is the case for continuous marginal distributions. We will not give the proof, but the result is given in Theorem 4.1.

Theorem 4.1 (Sklar)**.** *Let F be an n-dimensional distribution function with continuous margins $F_1, \cdots, F_n$. Then $\exists!$ $C : [0,1]^n \to [0,1]$ such that*

$$F(x_1, x_2, \cdots, x_n) = C(F_{x_1}(x_1), F_{x_2}(x_2), \cdots, F_{x_n}(x_n)) \quad \forall x \in \mathbb{R}.$$

Conversely, we also have that if $F_{x_1}, \cdots, F_{x_n}$ are distribution functions and C is an n-dimensional copula, then F is a distribution function and its margins are given by $F_{x_1}, \cdots, F_{x_n}$.

When one or more of the marginal distributions are not continous, this is no longer the case, see Genest and Neslehova (2007). We will continue with more properties of copulas, in particular for the 2-dimensional case. Schweizer and Wolff (1981) show that a copula is Lipschitz continuous and this gives us Proposition 4.2.

Proposition 4.2. *Any copula $C : [0,1]^n \to [0,1]$ is Lipschitz continuous with upper bound $\|C(u) - C(v)\| \leq \|u - v\|_1$ $\forall u, v \in [0,1]^n$.*

Proof. We will show it only for the bivariate case, but the proof for the n-dimensional case is similar.
Let us assume $0 \leq u_2 \leq v_2 \leq 1$. Then from property (3) of Definition 4.1 we obtain $C(u_1, v_2) - C(u_1, u_2) \leq C(1, v_2) - C(1, u_2) \leq v_2 - u_2$.
Then we generally have that $|C(u_1, v_2) - C(u_1, u_2)| \leq |v_2 - u_2|$ $\forall u_1, u_2, v_2 \in [0,1]$.
Similarly $|C(u_1, u_2) - C(v_1, u_2) \leq |v_1 - u_1|$ $\forall u_1, u_2, v_1 \in [0,1]$ and it follows that for general $u_1, u_2, v_1, v_2 \in [0,1]$:

$$\begin{aligned}
&|C(u_1, u_2) - C(v_1, v_2)| \\
&= \frac{1}{2}\{|C(u_1, u_2) - C(u_1, v_2) + C(u_1, u_2) - C(v_1, u_2) + C(v_1, u_2) - C(v_1, v_2) + C(u_1, v_2) - C(v_1, v_2)|\} \\
&\leq \frac{1}{2}\{|C(u_1, u_2) - C(u_1, v_2)| + |C(u_1, u_2) - C(v_1, u_2)| + |C(v_1, u_2) - C(v_1, v_2)| + |C(u_1, v_2) - C(v_1, v_2)|\} \\
&\leq \frac{1}{2}\{|u_2 - v_2| + |u_1 - v_1| + |u_2 - v_2| + |u_1 - v_1|\} \\
&= |u_1 - v_1| + |u_2 - v_2| \\
&= \|u - v\|_1.
\end{aligned}$$

□

Note that a copula is therefore also absolutely continuous and continuous. It does not follow that it is continuously differentiable, although from the absolute continuity, it does follow that the copula is continuously differentiable almost everywhere, see Aliprantis and Burkinshaw (1998).

For describing multivariate financial random variables we start working with the 2-dimensional copula and use this as a building block for higher dimensions. Filling in $n = 2$ in Definition 4.1 gives us the bivariate copula definition. This is given in Definition 4.2.

Definition 4.2. *A function $C : [0,1]^2 \to [0,1]$ is a 2-dimensional copula if and only if the following holds:*
(1) $\forall u \in [0,1]: \; C(0,u) = C(u,0) = 0$,
(2) $\forall u \in [0,1]: \; C(1,u) = C(u,1) = u$,
(3) $\forall (u_1,u_2)^T, (v_1,v_2)^T$, *with* $u_1 \le v_1$, $u_2 \le v_2$: $C(v_1,v_2) + C(u_1,u_2) \ge C(u_1,v_2) + C(v_1,u_2)$.

Example 4.1. *Some special cases of copulas are given below.*

(1) $C^0(u_1,u_2,\cdots,u_n) = u_1 u_2 \cdots u_n$.
(2) $C^+(u_1,u_2,\cdots,u_n) = \min\{u_1,\cdots,u_n\}$.
(3) $C^-(u_1,u_2,\cdots,u_n) = \max\{\sum_{i=1}^n u_i + 1 - n, 0\}$.

Here, C^0 is the independence copula describing the behaviour of independent distributions. This form is what we would expect, since filling in $F_{x_i}(x_i)$ for u_i gives us that:

$$F(x_1,x_2,\cdots,x_n) = C^0(F_{x_1}(x_1),F_{x_2}(x_2),\cdots,F_{x_n}(x_n)) = F_{x_1}(x_1)F_{x_2}(x_2)\cdots F_{x_n}(x_n),$$

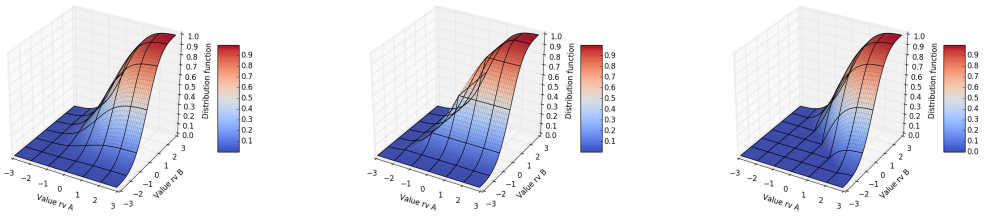
which must hold for independent data.

C^+ is the comonotonic copula, describing the behaviour of perfectly correlated distributions and C^- is the countermonotonic copula for $n = 2$, describing the behaviour of perfectly anti-correlated data. In higher dimensions this is not a copula, since it is not n-increasing for $n > 2$. This means that one of the marginal distributions is not non-decreasing (i.e. the density function is negative on a set $A \subset \mathbb{R}^n$ with measure greater than 0) and therefore this copula is not a distribution function.

The copulas of Example 4.1 for the 2-dimensional setting are shown in Figure 4.4. When the underlying distributions would be a standard normal distributions, these would correspond to a multivariate standard normal with $\rho = 0$, $\rho = 1$ and $\rho = -1$. We give these 2-dimensional distribution functions in Figure 4.3.

Figure 4.3: **The cases of copulas.**

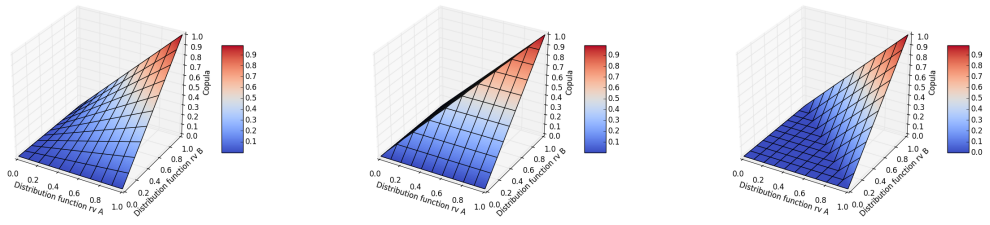
This figure shows three bivariate normal distributions with standard normal margins. The first figure shows the bivariate case with $\rho = 0$, the second $\rho = 1$ and the third $\rho = -1$.



(a) $\rho = 0$ (b) $\rho = 1$ (c) $\rho = -1$

Figure 4.4: **The three special cases of copulas.**

This figure shows the independent, comonotonic, and countermonotonic copula for the 2-dimensional setting.



(a) Independent copula (b) Comonotonic copula (c) Countermonotonic copula

Definition 4.3. *A function C_1 is bounded from below by a function C_2 on D (i.e. $C_1 \succ C_2$) if and only if $C_1(u) \geq C_2(u)\ \forall u \in D$.*
Similarly a function C_1 is bounded from above by a function C_2 on D (i.e. $C_1 \prec C_2$) if and only if $C_1(u) \leq C_2(u)\ \forall u \in D$ and obviously: $C_2 \succ C_1 \iff C_1 \prec C_2$.

Fréchet and Hoeffding showed independently for any copula that the counter- and comonotonic copula are a lower and upper bound respectively, see Nelsen (1999). This leads to Definition 4.3.

Theorem 4.3 (Fréchet-Hoeffding)**.** *For any copula $C : [0,1]^n \to [0,1]$ we have that $C^- \prec C \prec C^+$.*

Proof. $C \prec C^+$ follows directly from property (2) in Definition 4.1.
Furthermore, we know that C is Lipschitz continuous with bound $\|C(u) - C(v)\| = \|u - v\|_1 = \sum_{i=1}^n |u_i - v_i|$.
Thus, we have that for arbitrary $u, v \in [0,1]^n$: $C(u) = 1 - \|C(1,\cdots,1) - C(u)\| \geq 1 - \sum_{i=1}^n |1 - u_i| = 1 - n + \sum_{i=1}^n u_i$
and so $C^- \prec C$. □

In Propositions 4.4 and 4.5 we give some important properties of copulas. For details and proofs, that are straightforward from the definition of a copula, we refer to Nelsen (1999).

Proposition 4.4 (Survival Copula)**.** *Suppose C is a 2-dimensional copula for random variables X_1 and X_2. Then:*
(1) $\overline{C}$ defined by $\overline{C}(u_1,u_2) = u_1 + u_2 - 1 + C(1-u_1, 1-u_2)$ is also a copula.
(2) C_1 defined by $C_1(u_1,u_2) = u_2 - C(1-u_1, u_2)$ is also a copula.
(3) C_2 defined by $C_2(u_1,u_2) = u_1 - C(u_1, 1-u_2)$ is also a copula.
$\overline{C}$ is the so-called survival copula of C and C_i is the X_i-survival copula of C.

Proposition 4.5. *Suppose C_1 and C_2 are n-dimensional copulas. Then C defined by $C(u) := \lambda C_1(u) + (1-\lambda)C_2(u)\ \forall u \in [0,1]^n$ gives another copula if $\lambda \in [0,1]$. (The convex combination of two copulas is a copula).*

Corollary 4.6. *Directly from these two propositions we get that any convex combination of copulas and survival copulas is still a copula.*

Important properties of copulas are its dependence measures, that will be introduced in the next section. For extreme risk management we are especially interested in the tail dependence, which is a measure for the probability of extreme events for two random variables occurring at the same time. We are not only interested in the dependence in, but also outside the tails. In the next section we discuss a number of dependence measures for copulas.

4.3. Dependence measures for random variables

4.3.1. Tail dependence

In financial crises we see that most financial variables tend to move in the same direction and thus the dependence between variables is higher in the lower tail of the distribution, see Aloui et al. (2011). This extra dependence in the lower tail of the distributions suggests that there might be lower tail dependence. Similarly

there can be upper tail dependence and so it is important to come up with measures for this special dependence. We introduce the tail dependence coefficients (TDCs) as a measure for this and define it in Definition 4.4. We refer to Nelsen (1999) for this definition.

Definition 4.4. *Let X and Y be random variables with distribution functions F_x and F_y respectively. Then the upper tail dependence coefficient of $(X,Y)^T$ is given by:*

$$\lambda_U = \lim_{u\uparrow 1} \mathbb{P}\left(Y > F_y^{-1}(u) | X > F_x^{-1}(u)\right), \tag{4.3}$$

similarly the lower tail dependence coefficient is given by:

$$\lambda_L = \lim_{u\downarrow 0} \mathbb{P}\left(Y \leq F_y^{-1}(u) | X \leq F_x^{-1}(u)\right). \tag{4.4}$$

Definition 4.5. *We say that X and Y are upper (lower) tail dependent if $\lambda_U > 0$ ($\lambda_L > 0$). Similarly X and Y are upper (lower) tail independent if $\lambda_U = 0$ ($\lambda_L = 0$).*

From a risk management perspective we can also express the tail dependence in terms of VaR (Value at Risk), recall Definition 2.1. This expression is given in Proposition 4.7.

Proposition 4.7. *The coefficient of upper and lower tail dependence of random variables X and Y can be written as:*

$$\lambda_U = \lim_{u\uparrow 1} \mathbb{P}\left(Y > \mathrm{VaR}_u(Y) | X > \mathrm{VaR}_u(X)\right), \tag{4.5}$$

$$\lambda_L = \lim_{u\downarrow 0} \mathbb{P}\left(Y \leq \mathrm{VaR}_u(Y) | X \leq \mathrm{VaR}_u(X)\right). \tag{4.6}$$

A risk manager's interpretation for the lower tail dependence λ_L between two portfolios X and Y can be as follows. For small u we have: when a portfolio X drops below the VaR_u level (i.e. an extremely negative event for X) the probability of Y dropping below the VaR_u (i.e. an extremely negative event for Y) is approximately λ_L. Investing in two portfolios which have a large λ_L means that the probability of both dropping below the VaR_u level is relatively large.

Nelsen (1999) state that we can rewrite this definition for the tail dependence in a different manner and express it in terms of copulas. This expression is given in Proposition 4.8.

Proposition 4.8. *Let X and Y be random variables whose dependence structure is given by copula C (i.e. $F(X,Y) = C(F_x(X), F_y(Y))$). Then for the upper and lower tail dependence we have that:*

$$\lambda_U(C) = \lim_{u\uparrow 1} \frac{1 - 2u + C(u,u)}{1-u}, \tag{4.7}$$

$$\lambda_L(C) = \lim_{u\downarrow 0} \frac{C(u,u)}{u}. \tag{4.8}$$

These measures are important to model the dependence in extreme market conditions. Some financial variables tend to move in an opposite direction and therefore show no upper or lower tail dependence. Gold for example is known to increase in value during crises, see Baur and McDermott (2010), and to capture this in our copula we introduce another measure of tail dependence in Definition 4.6.

Definition 4.6. *Let X and Y be random variables whose dependence structure is given by copula C (i.e. $F(X,Y) = C(F_x(X), F_y(Y))$), then we can define the upper-lower tail dependence coefficient and the lower-upper tail dependence coefficient to be:*

$$\lambda_{U,L} = \lim_{u\downarrow 0} \mathbb{P}\left(Y > G^{-1}(1-u) | X \leq F^{-1}(u)\right) = 1 - \lim_{u\downarrow 0} \frac{C(1-u,u)}{u}, \tag{4.9}$$

$$\lambda_{L,U} = \lim_{u\downarrow 0} \mathbb{P}\left(Y \leq G^{-1}(u) | X > F^{-1}(1-u)\right) = 1 - \lim_{u\downarrow 0} \frac{C(u,1-u)}{u}. \tag{4.10}$$

Calculating the tail dependence coefficients for $C^-(u_1,u_2) = \max\{u_1+u_2-1,0\}$, $C^0(u_1,u_2) = u_1u_2$, $C^+(u_1,u_2) = \min\{u_1,u_2\}$ and some other widely used copulas gives us Table 4.1.

Table 4.1: **Tail dependence for important copulas**

This table gives the coefficients for tail dependence for seven copula families and the countermonotonic, comonotonic and independence copula.

Copula	λ_L	λ_U	$\lambda_{L,U}$	$\lambda_{U,L}$
C^-	0	0	1	1
C^0	0	0	0	0
C^+	1	1	0	0
Gaussian	0	0	0	0
Gumbel	0	$2-2^{\frac{1}{\alpha}}$	0	0
Clayton	$2^{\frac{1}{\alpha}}$	0	0	0
Frank	0	0	0	0
Plackett	0	0	0	0
Joe	$2^{\frac{1}{\alpha}}$	0	0	0
Student's t	$2-2t_\nu\left(\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$		0	0

As we can see from the table for the independence copula we indeed have that all the tail dependence measures are zero. The countermonotonic copula has no upper and lower tail dependence, but it is fully upper-lower and lower-upper tail dependent. The comonotonic copula is fully upper and lower tail dependent and therefore not upper-lower and lower-upper tail dependent. The Gumbel, Clayton, Joe and Student's t copula show tail dependence in one tail of the distribution. The Student's t copula shows tail dependence and is symmetric and thus has $\lambda_L = \lambda_U > 0$.

These coefficients only describe the dependence in the tails of the distribution but do not describe the dependence outside the tails. It is also important to come up with dependence measures for the whole range of possible outcomes, that will be the topic of the next section.

4.3.2. Other dependence measures

Widely used measures for the dependence between two variables over all possible outcomes is the Pearson's correlation coefficient and it is defined below.

Definition 4.7. *For X_1 and X_2 random variables with $\mathbb{E}\left[X_1^2\right], \mathbb{E}\left[X_2^2\right] < \infty$ the Pearson's correlation coefficient is given by:*

$$\rho = \rho(X_1,X_2) = \frac{\text{Cov}[X_1,X_2]}{\sqrt{(\text{Var}[X_1]\text{Var}[X_2])}}. \tag{4.11}$$

As discussed in Embrechts et al. (2003) there are some drawbacks in the use of this dependence coefficient, namely that:
(1) ρ depends on the marginal distributions and not only on the copula.
(2) ρ only exists when $\text{Var}[X_1]$ and $\text{Var}[X_2]$ exist.
(3) It is not invariant with respect to increasing transformations of X_1 and X_2 (i.e. $\rho(T_1(X_1),T_2(X_2)) \neq \rho(X_1,X_2)$ for increasing functions T_1 and T_2).

To avoid having these drawbacks we introduce two other important measures of dependence, namely Kendall's tau and Spearman's rho. Those two measures are alternatives to a linear correlation coefficient which may be inappropriate. To define them we use the number of concordant and discordant pairs. Whether a pair is concordant (move in the same direction) or discordant (move in opposite directions) depends on the rank of the data. This concordance concept is defined below.

Definition 4.8. *A pair $\{X_1^i, X_2^i\}$ and $\left\{X_1^j, X_2^j\right\}$ is concordant if and only if $\left(R_1^i - R_1^j\right)\left(R_2^i - R_2^j\right) > 0$ and discordant if and only if $\left(R_1^i - R_1^j\right)\left(R_2^i - R_2^j\right) < 0$.*

Then Kendall's tau is simply the probability of concordance minus the probability of discordance and it is defined below.

Definition 4.9. *We define Kendall's tau for a random vector* $(X,Y)^T$ *to be equal to:*

$$\tau(X,Y) = \mathbb{P}\left((X-\tilde{X})(Y-\tilde{Y})>0\right) - \mathbb{P}\left((X-\tilde{X})(Y-\tilde{Y})<0\right), \tag{4.12}$$

where $(\tilde{X},\tilde{Y})^T$ *is an independent copy of* $(X,Y)^T$*. Spearman's rho is similar and defined below.*

Definition 4.10. *We define Spearman's rho for a random vector* $(X,Y)^T$ *to be equal to:*

$$\rho_S(X,Y) = 3\left(\mathbb{P}\left((X-\tilde{X})(Y-Y')>0\right) - \mathbb{P}\left((X-\tilde{X})(Y-Y')<0\right)\right), \tag{4.13}$$

where $(\tilde{X},\tilde{Y})^T$ *and* $(X',Y')^T$ *are independent copies of* $(X,Y)^T$ *(also mutually independent).*

In Propositions 4.9 and 4.10 we show that we can express those dependence measures as functions of the ranks only.

Proposition 4.9. *Let us have* m *observations* $\{X^1,\cdots,X^m\}$ *where* $X^i = (X_1^i, X_2^i)^T$ *and let* c *be the number of concordant and* d *be the number of discordant pairs. Then we have that:*

$$\tau = \frac{2(c-d)}{m(m-1)}. \tag{4.14}$$

Proposition 4.10. *Let us have* m *observations* $\{X^1,\cdots,X^m\}$ *where* $X^i = (X_1^i, X_2^i)^T$ *and let* R_1^i *and* R_2^i *be the corresponding ranks. Then we have that:*

$$\rho_S = 1 - \frac{6\sum\left(R_1^i - R_2^i\right)^2}{m(m^2-1)}. \tag{4.15}$$

These two measures do not suffer from the three drawbacks of the Pearson's rho. There is, however, a new drawback here, which occurs when we have ties in our observations. Under the assumption of continuous density functions, in theory, two or more arbitrary observations are unequal almost surely. In practice on the other hand, they can be equal due to the truncation of the data. There are several ways to deal with this, since the rank is not uniquely defined in this case. In Appendix E we explain the different possible ranking methods and the combination of ordinal and the fractional ranking method. This ranking choice solves for the non-uniqueness problem and as a result we obtain two robust dependence measures.

The ρ_S and τ that follow from our ranks and ranking method are important characteristics of a copula. We can directly express the measures ρ_S and τ in terms of copulas as well, which is shown in Nelsen (1999). The results are shown in the Proposition 4.11.

Proposition 4.11. *Let* X *and* Y *be random variables whose dependence structure is given by copula* C *(i.e.* $F(X,Y) = C(F_x(X), F_y(Y))$*). Then for Kendall's tau we have that:*

$$\tau(X,Y) = 4\int\int_{[0,1]^2} C(u,v)\,dC(u,v) - 1 = 4\mathbb{E}[C(U,V)] - 1, \tag{4.16}$$

where $U,V \sim U(0,1)$ *with joint distribution* C*. For Spearman's rho we have that:*

$$\rho_S(X,Y) = 12\int\int_{[0,1]^2} uv\,dC(u,v) - 3 = 12\int\int_{[0,1]^2} C(u,v)\,du\,dv - 3. \tag{4.17}$$

From this proposition we see that these two measures do not depend on the marginal distributions, but only on the copula. Directly from the definitions we see that ρ_S and τ always exist. Also these two measures are invariant with respect to an increasing transformation, since they can be expressed in term of a copula, which itself is invariant with respect to increasing transformations of its margins. Thus, none of the drawbacks that apply for ρ apply for ρ_S and τ.

4.3.3. Link between dependence measures and copulas

In the previous sections we saw that we can express the τ, ρ_S and TDCs in terms of copulas. Thus, for a survival copula we can express all these dependence measures in terms of the dependence measures of the original copula. These expressions are given in Nelsen (1999) and in Proposition 4.12. Similarly we can link the dependence measures of a linear combination of copulas C_1 and C_2 to the individual dependence measures of C_1 and C_2, that are given in Proposition 4.13.

Proposition 4.12. *Suppose we have a copula C with dependencies $\lambda_L(C), \lambda_U(C), \lambda_{U,L}(C), \lambda_{L,U}(C), \tau(C)$ and $\rho_S(C)$. Then the following holds.*
For the survival copula $\overline{C}$: $\lambda_L(\overline{C}) = \lambda_U(C)$ and $\lambda_U(\overline{C}) = \lambda_L(C)$ (i.e. the survival copula's upper tail dependence is equal to the copula's lower tail dependence and vice versa).
Similarly we have that $\lambda_{L,U}(C) = \lambda_{U,L}(\overline{C})$. Furthermore, $\tau(\overline{C}) = \tau(C)$, $\rho_S(\overline{C}) = \rho_S(C)$ (i.e. the Kendall's tau and Spearman's rho of the survival copula are equal to those of the original copula).

Proposition 4.13. *Assume we have copulas C_i for $i = 1,2$ with dependencies $\lambda_L(C_i), \lambda_U(C_i), \tau(C_i)$ and $\rho_S(C_i)$. Then the following holds for $C = \alpha C_1 + (1-\alpha)C_2$, with $\alpha \in [0,1]$:*
(1) $\lambda(C) = \alpha\lambda(C_1) + (1-\alpha)\lambda(C_2)$ for all four tail dependence coefficients.
(2) $\rho_S(C) = \alpha\rho_S(C_1) + (1-\alpha)\rho_S(C_2)$.
(3) $\tau(C) = \alpha\tau(C_1) + (1-\alpha)\tau(C_2)$ does not necessarily hold.

Proof. (1) follows from the linearity of the limit and (2) from the linearity of the integral. We show this for (2) only:

$$\begin{aligned}
\rho_S(C) &= 12\int\int_{[0,1]^2} C(u,v)dudv - 3 \\
&= 12\left(\alpha\int\int_{[0,1]^2} C_1(u,v)dudv + (1-\alpha)\int\int_{[0,1]^2} C_2(u,v)dudv\right) - 3 \\
&= \alpha\left(12\int\int_{[0,1]^2} C(u,v)dudv - 3\right) + (1-\alpha)\left(12\int\int_{[0,1]^2} C_2(u,v)dudv - 3\right) \\
&= \alpha\rho_S(C_1) + (1-\alpha)\rho_S(C_2).
\end{aligned}$$

For (3) we give a counterexample. Let $C_1(u,v) = C^0(u,v) = uv$ and $C_2(u.v) = C_+(u,v) = \min\{u,v\}$. Then for $C(u,v) = \frac{1}{2}(C_1(u,v) + C_2(u,v))$ we obtain the following:

$$\begin{aligned}
\tau(C) &= 4\int\int_{[0,1]^2} C(u,v)c(u,v)dudv - 1 \\
&= 4\frac{1}{4}\int\int_{[0,1]^2} (C_1(u,v) + C_2(u,v))(c_1(u,v) + c_2(u,v))dudv - 1 \\
&= \int\int_{[0,1]^2} (uv + \min\{u,v\})(1 + \delta(u-v))dudv - 1 \\
&= \int\int_{[0,1]^2} uvdudv + \int\int_{[0,1]^2} uv\delta(u-v)dudv + \int\int_{[0,1]^2} \min\{u,v\}dudv \\
&\qquad + \int\int_{[0,1]^2} \min\{u,v\}\delta(u-v)dudv - 1 \\
&= \frac{1}{4} + \frac{1}{2} + \frac{1}{3} + \int_0^1 \left(\int_0^v udu + \int_v^1 vdu\right)dv - 1 \\
&= \frac{1}{4} + \frac{1}{2} + \frac{1}{3} + \left(\frac{1}{2} - \frac{1}{6}\right) - 1 = \frac{5}{12} \\
&\neq \frac{1}{2}\tau(C_1) + \frac{1}{2}\tau(C_2).
\end{aligned}$$

□

4.4. COPAR models

Copula theory is a valuable theory that we can use to model tail dependence. To model tail dependence between factors we can implement copulas in the modelling for the statistical factors. Recall that the factors

are currently modelled with a VAR model where the errors are normally distributed and show dependence but no tail dependence. To be able to model the tail dependence between the factors we can use copulas to describe the dependence in the errors. Describing the dependence in the errors by a copula generalizes our VAR model to a copula autoregressive (COPAR) model. We formally define this COPAR model in Definiton 4.11.

Definition 4.11 (COPAR model)**.** *A copula autoregressive model of order p (COPAR(p) model) is of the form:*

$$y^{(t)} = c + \sum_{i=1}^{p} A_i y^{(t-i)} + \epsilon^{(t)}, \tag{4.18}$$

where $c \in \mathbb{R}^n$ and $A_i \in \mathbb{R}^{n\times n}$ are constants $\forall\, i$, A_p has at least one non-zero element and $\epsilon^{(t)} \in \mathbb{R}^n$ is time independent, zero-mean and its dependence can be described by a copula C.

If the errors exhibit correlation due to the copula then the factors will exhibit correlation as well. Also tail dependence in the errors implies tail dependence in the factors. We can unfortunately not directly link the TDCs of the errors to the TDCs of the factors, since this depends on the matrices A_i and possibly on underlying distributions as well. Larger tail dependence in the errors will however imply larger tail dependence in the factors. We will do some simulations in section 5.5 to show that there is a positive relation between the TDCs of the errors and TDCs of the factors in our test scenarios.

4.5. Parametric copulas

In many applications parametric copulas can be used to model the dependence between random variables. A special case of parametric copulas are the Archimedean copulas. This class of copulas is convenient to use, because of its simple mathematical form, which we will elaborate on later. The Gaussian copula is an example of an Archimedean copula and is simply a copula which describes multivariate normal distributions. Li (2000) shows how default correlation can be modelled using these copulas. After this introduction some years before the credit crisis, the Gaussian copula was widely used to estimate especially default probabilities. Salmon (2012) discuss this and state that the use of this model could have contributed to the credit crisis. The Gaussian method was adopted by many investors, yet it has drawbacks that were not always considered. The Gaussian copula namely does not include tail dependence (unless $\rho = 1$) and it can therefore underestimate the risk if data shows lower tail dependence. Malevergne and Sornette (2003) address this problem and show that the use of the Gaussian copula can be inappropriate in financial settings. The formula for the Gaussian copula is given by:

$$C^N(u_1, u_2) = \Phi_R(\Phi^{-1}(u_1) + \Phi^{-1}(u_2)), \tag{4.19}$$

with Φ the cumulative distribution function of a standard normal and Φ_R the joint distribution function of a multivariate zero-mean normal with covariance matrix R (equal to the correlation matrix). This copula can capture the dependence in the centre of the distribution, but due to the Gaussian form all the variables are tail independent, unless $\rho = 1$. We know that financial variables can show large positive correlation in a financial crisis even when they usually show small or negative correlation. This suggests that there might be lower tail dependence in the data and we should be able to capture this with our copula, which cannot be done with the Gaussian copula. This is an argument against the use of the Gaussian copula to model dependence of financial variables.

When we wish to model financial data using a parametric copula we can link the parameters of the model to the Kendall's tau, Spearman's rho, and/or tail dependence coefficients. Estimations of those measures consequently give information on what parametric copula we can use and what the values of the parameters in the model should be. Using linear combinations of copulas and survival copulas give us freedom in our parameters, such that we can make sure that the copula we use has the same dependence coefficients as the data.

The Gaussian copula introduced in the previous section is a special case of an Archimedean copula. The popularity of Archimedean copulas comes from their simple mathematical form. Archimedean copulas are copulas generated by a function g and depends on the pseudo-inverse $g^{[-1]}$ as well, which is defined as follows.

Definition 4.12. *Let $g : [0,1] \to [0,\infty]$ be a continuous, strictly decreasing function with $g(1) = 0$. Then the*

pseudo-inverse of g is $g^{[-1]}:[0,\infty]\to[0,1]$ and given by:

$$g^{[-1]}(t)=\begin{cases}g^{-1}(t), & 0\le t\le g(0),\\ 0, & g(0)\le t\le\infty.\end{cases} \tag{4.20}$$

Here, $g(0)=\infty\Longleftrightarrow g^{[-1]}=g^{-1}$ (i.e. the pseudo-inverse of g equals the inverse of g).

We can now define the Archimedean copula as follows.

Definition 4.13. *A copula $C:[0,1]^n\to[0,1]$ is called an Archimedean copula if and only if there exists a continuous, strictly decreasing, convex g such that:*

$$C(u_1,\cdots,u_n)=g^{[-1]}(g(u_1)+\cdots+g(u_n)),$$

where g is called the generator of C. If $g^{[-1]}=g^{-1}$, g is called a strict generator and C the corresponding strict Archimedean copula.

This expression in terms of a generator is convenient since it gives a simple representation of a copula in n dimensions. In addition, we can now express the Kendall's tau in terms of the generator using equation (4.16) and simplify, see Nelsen (1999), to obtain:

$$\tau=1+4\int_0^1\frac{g(t)}{g'(t)}dt. \tag{4.21}$$

Using equation (4.17) we can also compute the Spearman's rho as a function of g, this can generally not be simplified and gives:

$$\rho_S=12\int\int_{[0,1]^2}g^{-1}(g(u_1)+g(u_2))du_1du_2-3. \tag{4.22}$$

We proceed with some examples of widely used Archimedean copulas. Nelsen (1999) summarizes a list of many copulas and we give five below.

Example 4.2. *Some examples of 2-dimensional Archimedean copulas are:*

$$\begin{aligned}
&C^A(u_1,u_2;\theta)=\tfrac{u_1u_2}{1-\theta(1-u_1)(1-u_2)}, && \theta\in[-1,1], && \text{(Ali-Mikhail-Haq copula)},\\
&C^C(u_1,u_2;\theta)=\max\{(u_1^{-\theta}+u_2^{-\theta}-1)^{-\theta^{-1}},0\}, && \theta\in[-1,\infty]\setminus\{0\}, && \text{(Clayton copula)},\\
&C^F(u_1,u_2;\theta)=-\tfrac{1}{\theta}\log\left[1+\tfrac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right], && \theta\neq0, && \text{(Frank copula)},\\
&C^G(u_1,u_2;\theta)=\exp\left[-\left\{(-\log u_1)^\theta+(-\log u_2)^\theta\right\}^{\theta^{-1}}\right], && \theta\in[1,\infty], && \text{(Gumbel copula)},\\
&C^J(u_1,u_2;\theta)=1-\left[(1-u_1)^\theta+(1-u_2)^\theta-(1-u_1)^\theta(1-u_2)^\theta\right]^{\theta^{-1}}, && \theta\in[1,\infty], && \text{(Joe copula)}.
\end{aligned}$$

If we wish to fit a certain model to the data, we can consider some copulas and linear combinations of copulas and estimate the parameter(s) in the model. The choice of the final copula can then be made by comparison of for example the Akaike information criterion, see Akaike (1974). There are different methods to estimate the parameters in these models. In section E.2 in the appendix we will discuss how we can estimate the parameters using maximum likelihood estimation (MLE).

4.5.1. Pros and cons parametric copulas

It is convenient to assume some parametric form for copulas, especially when they take a simple mathematical form, like Archimedean copulas. Archimedean copulas are convenient for higher dimensional cases too, because of the simple form and associativity. We can write them as follows:

$$C(u_i,C(u_1,\cdots,u_{i-1},u_{i+1},\cdots,u_n))=C(u_1,\cdots,u_n)\ \forall i. \tag{4.23}$$

Also we can use MLE to find the parameters in the model and use known sampling methods to generate scenarios.

However, we should be careful in doing so, since assuming a certain form goes hand in hand with excluding certain behaviour. For example assuming a Gaussian copula excludes tail dependence, while Aloui

et al. (2011) show there are indications of tail dependence. Using a copula which includes tail dependence might not appropriately describe the dependence in other parts of the distribution. All these Archimedean copulas have only one or two parameters, which might accurately describe 2-dimensional behaviour, but this accuracy does not necessarily carry over to higher dimensions. We could also choose to take a convex combination of copulas and survival copulas to be more flexible, but this makes it more difficult to estimate the parameters, since there are more. Furthermore, it is more difficult to sample from them. Also a linear combination of Archimedean copulas is not necessarily an Archimedean copula, which is inconvenient for especially higher dimensional cases, since we lose the associativity. Hence, we preferably do not assume any parametric form and introduce semiparametric copulas.

4.6. Semiparametric copulas and tail dependence estimation

In a semiparametric copula approach we do not assume any parametric form of the copula. However, we make an assumption by choosing a certain approach and if we do not take care, we may exclude specific behaviour here as well. We first introduce the empirical copula, which is based on the same concept as the empirical distribution function. We will show later that this empirical copula is strictly speaking not a copula.

Definition 4.14. *Let $\{X^1, \cdots, X^m\} \in \mathbb{R}^n$ be independent and identically distributed random variables from distribution function F. Then its empirical distribution function is given by:*

$$\hat{F}(x_1, \cdots, x_n) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}\{X_1^i \leq x_1, \cdots, X_n^i \leq x_n\}. \tag{4.24}$$

Since we also have that $X_j^i \leq x_j \iff F_{x_j}^{-1}\left(X_j^i\right) \leq F_{x_j}^{-1}(x_j)$ and $F(x_1, \cdots, x_n) = C\left(F_{x_1}^{-1}(x_1), \cdots, F_{x_n}^{-1}(x_n)\right)$, we can get an estimation for the copula. This so-called empirical copula is defined in Definition 4.15.

Definition 4.15. *Let $\{X^1, \cdots, X^m\} \in \mathbb{R}^n$ be independent, identically distributed random variables from distribution function F. Then the corresponding empirical copula, of which we will show that it is not a copula later, is given by:*

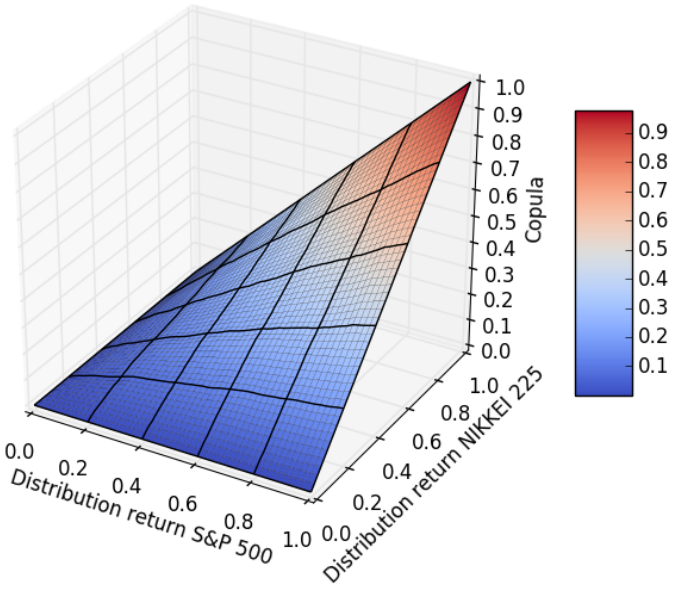
$$\hat{C}_n(u_1, \cdots, u_n) = \frac{1}{m} \sum_{j=1}^{m} \mathbb{1}\left\{ \frac{R_1^j}{m} \leq u_1, \cdots, \frac{R_n^j}{m} \leq u_n \right\}, \tag{4.25}$$

where $1 \leq R_i^j \leq m$ denotes the rank of X_i^j with respect to $X_i^1, \cdots X_i^m$ $\forall 1 \leq i \leq n$.

We perform this estimation for the copula between the Tokyo and the New York stock exchange indices (i.e. the Nikkei 225 and S&P 500). The copula is shown in Figure 4.5.

Figure 4.5: **The empirical copula for the Nikkei 225 and S&P 500 for the years 2006-2016.**

This figure shows the dependence structure between the daily log-return of the Nikkei 225 and S&P 500 for the period between 2006 and 2016 .



Note that this can be done under the assumption that the multivariate distribution function is static, which does not need to be the case here. In Chapter 5 we will make this assumption for the underlying errors in the dynamic factor modelling, details can be found in that section.

This empirical copula is not smooth, since the indicator function causes it to jump. From Proposition 4.2, which states that a copula is Lipschitz continuous, it must then also be continuous, thus the empirical copula is not a copula. Based on this copula estimation formula we can estimate the TDC for the 2-dimensional case. Recall equations (4.7)-(4.10), that link the copula to tail dependence and enables us to obtain estimators for the TDCs. Schmidt (2005) shows this and we obtain the following.

$$\begin{aligned}
\hat{\lambda}_U^{(1)}(k) &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\{R_1^i > n-k, R_2^i > n-k\},\\
\hat{\lambda}_L^{(1)}(k) &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\{R_1^i \le k, R_2^i \le k\},\\
\hat{\lambda}_{U,L}^{(1)}(k) &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\{R_1^i > n-k, R_2^i \le k\},\\
\hat{\lambda}_{L,U}^{(1)}(k) &= \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\{R_1^i \le k, R_2^i > n-k\},
\end{aligned} \tag{4.26}$$

where we should have that $k, \frac{n}{k} \to \infty$ for these to converge to the theoretical coefficients. We can also write:

$$\lambda_U = \lim_{v\uparrow 1} \frac{1-2v+C(v,v)}{1-v} = 2 - \lim_{v\uparrow 1} \frac{1-C(v,v)}{1-v} = 2 - \lim_{v\uparrow 1} \frac{\log C(v,v)}{\log v} \approx 2 - \frac{\log \hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)}{\log \frac{n-k}{n}},$$

where the approximation should be close for k small. Using this approximation and similar approximations

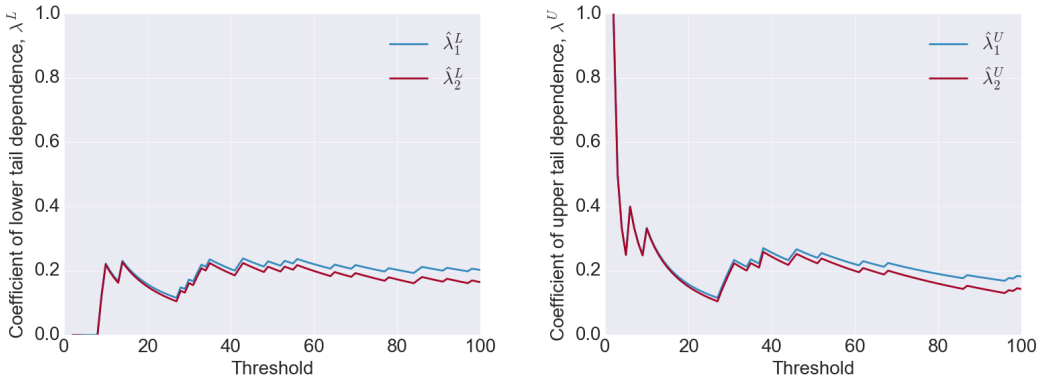
for λ_L, $\lambda_{U,L}$ and $\lambda_{L,U}$, introduced by Coles et al. (1999), we can also obtain the following estimators.

$$\begin{aligned}
\hat{\lambda}_U^{(2)} &= 2 - \frac{\log \hat{C}_n\left(\frac{n-k}{n}, \frac{n-k}{n}\right)}{\log \frac{n-k}{n}}, \\
\hat{\lambda}_L^{(2)} &= \frac{\log\left(1 - \hat{C}_n\left(\frac{k}{n}, \frac{k}{n}\right)\right)}{\log \frac{n-k}{n}}, \\
\hat{\lambda}_{L,U}^{(2)} &= 1 - \frac{\log\left(1 - \hat{C}_n\left(\frac{n-k}{n}, \frac{k}{n}\right)\right)}{\log \frac{n-k}{n}}, \\
\hat{\lambda}_{U,L}^{(2)} &= 1 - \frac{\log\left(1 - \hat{C}_n\left(\frac{k}{n}, \frac{n-k}{n}\right)\right)}{\log \frac{n-k}{n}}.
\end{aligned} \tag{4.27}$$

that are the same in the limit and therefore approximately the same for small k, where k is to be chosen as threshold. We have illustrated the estimates for the lower and upper tail dependence coefficients for the Nikkei 225 and the S&P 500 in Figure 4.6.

Figure 4.6: **The tail dependence coefficients for different thresholds**

This figure shows the three estimators for the coefficents of lower and upper tail dependence between the Nikkei 225 and the S&P 500.



(a) Lower tail (b) Upper tail

Both estimators are accurate when there is tail dependence in the data. However, they always give values greater than or equal to zero. For uncorrelated data, we can show that the expectation of this estimator $\mathbb{E}[\hat{\lambda}_i] > 0$, while we know that $\lambda_U = \lambda_L = 0$. There is namely always a possible probability that $\hat{\lambda}_i > 0$, unless the data has $\tau = -1$.

Thus, before using these estimates we wish to test if there even is tail dependence, which will be the topic of section 4.8. Before we proceed with this test, we will discuss ways to select a threshold and show that a regression method can be appropriate for determining λ in the next section.

4.7. Threshold selection for the tail dependence coefficients estimators

As we have seen in the previous section we still need to choose a threshold for both estimators. Similar to previous threshold methods, we are dealing with an increasing bias of our estimate and decreasing variance of our estimate with respect to the threshold. We therefore again need to make a bias efficiency trade-off. In this section we introduce two methods for dealing with this problem. The first method is one which looks for the first plateau where the estimate does not change that much, which is similar to the Stability method for the tail index. The other method uses regression of the estimates as a function of the threshold. In the latter method we base our estimate $\hat{\lambda}$ on $\hat{\lambda}(1), \cdots \hat{\lambda}(m)$, where we still have to choose m.

4.7.1. Plateau-finding method

Recall the Stability method introduced in section 2.4. We designed this method for the Hill estimator in particular, but it also worked for the other tail index estimators. In all these tail index estimators the bias efficiency trade-off needs to be made and this trade-off occurs here too. We can simply fill in the estimates $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ in Algorithm 2.2 instead of filling in $\hat{\gamma}^H$. The main difference between this data and the data for the tail index is that this data is never constant. Taking an extra value into account for the estimation of λ (i.e. increasing the threshold by one) always makes the estimator jump up or down, unless it is zero or one. Because of these jump that will always occur in tail dependent data with $\lambda < 1$, we can expect this method not to work as well as for the tail index estimations. Smoothing the estimates for determination of the threshold should improve this method and Frahm et al. (2005) use this so-called Plateau-finding method. We have considered both approaches and as expected the plateau method is better in MSE. Besides we come up with a regression method to determine the TDC's. The method and the idea behind the method are described in the next section.

4.7.2. Regression method

Instead of using one value $\hat{\lambda}(k)$ for $k \in \{1, \cdots, m\}$ we can also use more values $\hat{\lambda}(1), \cdots \hat{\lambda}(m)$, that all give information about λ. We know that the value of the estimator in $k = 1$ has the lowest bias but the highest variance. There is no bias in 0, but we do not have a value for the estimator in 0. If the bias increases linearly in the threshold, we can regress the estimates on the thresholds and then the value of the regressed line in 0 should have no bias. This should therefore consistently approximate the theoretical value of $\hat{\lambda}$. This linearity assumption does not generally hold, because the bias depends on the rate of convergence towards the TDC, which itself depends on the underlying copula as we can see in Proposition 4.8. This bias therefore does not need to be linear in the threshold. This linear approximation does however give an approximation of the TDC and in section 4.11 we will show that it performs better in MSE than the Plateau-finding method on the simulated data we considered (i.e. for Clayton and Gumbel copulas).

Since the variance is not constant we could also use a WLS regression, but we do not know which weights we should use. The relation between the variance of the estimate and the threshold depends on the underlying distribution and is therefore not known. We here consider a WLS regression with the square root of the thresholds as weights and a regression with equal weights and compare the results with the Plateau-finding method in section 4.11.

4.8. Hypothesis testing for tail dependence

Before using the above mentioned method, we first wish to check whether there even is tail dependence at all. We will do this via hypothesis testing, where as the null hypothesis we take H_0: "the data is tail dependent." We do this because from a risk management perspective it is worse to underestimate risk than to overestimate risk. Rejecting tail dependence under tail dependence corresponds to underestimating the risk. Conversely, not rejecting tail dependence under tail independence corresponds to overestimating the risk. We therefore wish to reject tail dependence under tail dependence only with small probability α, the so-called significance level. Equation (4.28) gives α mathematically, see Young and Smith (2005).

$$\alpha = \mathbb{P}\left(\text{reject } H_0 | H_0\right). \tag{4.28}$$

Another important quantity in hypothesis testing is the power p, which we give in equation(4.29), see Young and Smith (2005).

$$p = \mathbb{P}\left(\text{reject } H_0 | H_1\right) \tag{4.29}$$

We want our test to have p as large and α as small as possible. To do this we use a method introduced by Reiss and Frick (2009) which uses the radial component of the data. We define the radial component in Definition 4.16.

Definition 4.16 (Radial component)**.** *The radial components $T^1, \cdots, T^m$ for a sample $X^1, \cdots, X^m \in \mathbb{R}^n$ are given by $T^j = \sum_{i=1}^n X_i^j$ for $1 \leq j \leq m$.*

Theorem 4.14 shows that we can link the radial component to tail dependence.

Theorem 4.14. *For random variable $X_1, X_2, \cdots, X_n$ described by distribution functions $F_{x_1}, F_{x_2}, \cdots F_{x_n}$ we can define $U_i = F_{x_i}^{-1}(X_i) - 1 \forall i$. Then for the radial component T of the vector U we have that:*

$$\mathbb{P}\left(T > ct | T > c\right) = t^{1+\beta} \text{ for some } \beta \geq 0, \tag{4.30}$$

where $\beta = 0$ is equivalent to having upper tail dependence.

Proof. The proof of this theorem is given by Falk and Michel (2006) and we will not elaborate on this here. □

From a data set we can simply compute the radial components and use them to determine whether we are dealing with tail dependence or not. Reiss and Frick (2009) state that the Neyman-Pearson test is uniformly most powerful and its critical region as function of α is given in equation (4.31).

$$K_{m,\alpha} = \left\{ \sum_{i=1}^{m} \log \frac{T_{i,n}}{T_{m,n}} > H_m^{-1}(1-\alpha) \right\}, \tag{4.31}$$

where $T_{1,n}, \cdots, T_{n,n}$ are the ordered $T_1, \cdots T_n$ such that $T_{1,n}$ is the largest value. This critical region is an asymptotic result for $m \to \infty$ and $\frac{m}{n} \to 0$, so for finite sample sizes we can expect a larger percentage of incorrect rejections than α. We will use this test for different α and approximate the probability of rejection for these samples (divide the number of rejections by the amount of simulations). For $m \to \infty$ and $\frac{m}{n} \to 0$ we expect these percentages of tail dependent samples to converge to α.

The results of these simulations are given in section 4.11. Here, we considered samples from a Clayton and Gumbel copula to see if these approximated percentages indeed converge to the theoretical values for different TDCs. Next, we will consider samples from a Gaussian copula to see what the power of the test is for Gaussian copulas (i.e. probability of rejection under H_1).

4.9. Subcopulas and extension

In the section 4.6 we introduced the empirical copula, which contains discontinuities and is therefore not a copula. The lower TDC of the empirical copula is 0 and the upper tail dependence does not exist even if there is lower and/or upper tail dependence in the data. Since we wish to be able to include tail dependence in our copula we need a different approach. A possible method is via the extension of a subcopula, which will be the topic of this section. This subcopula is defined on a subset $D \subset [0,1]^n$, for example on a grid, and we expand this continuously to obtain a continuous copula on $[0,1]^n$. Using this approach we can make sure that the copula has the right properties on the subset. One of these properties is tail dependence and in section 4.9.1 we show that this method is appropriate for including tail dependence. Let us first consider the 2-dimensional case. Higher dimensional cases are discussed in section 4.9.4.

Definition 4.17. *$C_s : D_1 \times D_2 \to [0,1]$ is called a 2-dimensional subcopula if and only if the following properties hold.*
(1) D_1, D_2 are subsets of $[0,1]$, both containing 0 and 1.
(2) $C_s(u_1, 0) = C_s(0, u_2) = 0 \ \forall u_1 \in D_1$ and $u_2 \in D_2$.
(3) $C_s(u_1, 1) = u_1$ and $C_s(1, u_2) = u_2 \ \forall u_1 \in D_1$ and $u_2 \in D_2$.
(4) $C_s(u_1, u_2) + C_s(v_1, v_2) \geq C_s(v_1, u_2) + C_s(u_1, v_2) \ \forall u_1 \leq v_1$ and $u_2 \leq v_2$, where $u_1, v_1 \in D_1$ and $u_2, v_2 \in D_2$.

Theorem 4.15. *Let C_s be a subcopula defined on $D_1 \times D_2$. Then there exists a copula C such that:*

$$C(u_1, u_2) = C_s(u_1, u_2) \ \forall u_1 \in D_1, \ u_2 \in D_2.$$

(i.e. any subcopula can be extended to a copula.)
Note: this C is not necessarily unique.

Proof. The proof is given by Schweizer and Sklar (1974). □

They show that the subcopula C_s can be extended to a subcopula $\overline{C_s}$ on $\overline{D_1} \times \overline{D_2}$ the closure of $D_1 \times D_2$. For any point (u_1, u_2) not in the domain of $\overline{C_s}$ we can use a 2-dimensional linear interpolation and define:

$$\begin{aligned} C_L(u_1, u_2) = &(1-\lambda_1)(1-\mu_1)\overline{C_s}(u_1^-, u_2^-) + (1-\lambda_1)\mu_1\overline{C_s}(u_1^-, u_2^+) \\ &+ \lambda_1(1-\mu_1)\overline{C_s}(u_1^+, u_2^-) + \lambda_1\mu_1\overline{C_s}(u_1^+, u_2^+), \end{aligned} \tag{4.32}$$

where $u_1^-, u_1^+ \in \overline{D_1}$ are the points closest to u_1 such that $u_1^- \leq u_1 \leq u_1^+$ and $u_2^-, u_2^+ \in \overline{D_2}$ are the points closest to u_2 such that $u_2^- \leq u_2 \leq u_2^+$, λ_1 and μ_1 are defined as follows:

$$\lambda_1 = \begin{cases} \frac{u_1 - u_1^-}{u_1^+ - u_1^-}, & \text{if } u_1^- < u_1^+, \\ 1, & \text{if } u_1^- = u_1^+, \end{cases}$$

$$\mu_1 = \begin{cases} \frac{u_2 - u_2^-}{u_2^+ - u_2^-}, & \text{if } u_2^- < u_2^+, \\ 1, & \text{if } u_2^- = u_2^+. \end{cases}$$

This is always a copula and we call this the linear extension of our copula.

Suppose we have a set of 2-dimensional data with m observations, then we can estimate the copula on a grid (grid copula). Theorem 4.16 gives the estimate on this grid and Schweizer and Sklar (1974) state that this is a subcopula.

Theorem 4.16. *Let C_G be defined as follows:*

$$C_G\left(\frac{i}{m}, \frac{j}{m}\right) = C_G(a_i, b_j) = \frac{1}{m}\sum_{k=1}^{m} \mathbb{1}\left(R_1^k \le i, R_2^k \le j\right) \forall 0 \le i, j \le m, \tag{4.33}$$

Then this a subcopula.

Proof. From Definition 4.17 we see that the first three subcopula conditions are clearly satisfied and the last condition is also satisfied, since $\mathbb{1}\left(R_1^k \le i, R_2^k \le j\right) + \mathbb{1}\left(R_1^k \le a, R_2^k \le b\right) \ge \mathbb{1}\left(R_1^k \le a, R_2^k \le j\right) + \mathbb{1}\left(R_1^k \le i, R_2^k \le b\right)$ $\forall i \ge a,\ j \ge b$. □

The copula is defined on a grid which is a union of points and therefore closed. Since the closure of the grid is the grid itself, we can immediately use the extension defined by equation (4.32). We denote this extension by $C_{G,L}$, which describes the dependence in the data. To see if this copula has the desired properties, we first calculate the tail dependence coefficients.

Proposition 4.17. *The linear extension of the grid subcopula, defined by equation (4.33) has the following TDCs:*

$$\lambda_L = \lambda_U = \lambda_{L,U} = \lambda_{U,L} = 0$$

.

Proof.

$$\begin{aligned}
\lambda_L &= \lim_{a \downarrow 0} \frac{C(a,a)}{a} \\
&= \lim_{a \downarrow 0} \frac{1}{a}\left((1-\lambda_1)(1-\mu_1)C(0,0) + (1-\lambda_1)\mu_1 C(0,b_1) + \lambda_1(1-\mu_1)C(a_1,0) + \lambda_1\mu_1 C(a_1,b_1)\right) \\
&= \lim_{a \downarrow 0} \frac{1}{a}\lambda_1\mu_1 C(a_1,b_1) \\
&= \lim_{a \downarrow 0} \frac{1}{a}\frac{a}{a_1}\frac{a}{b_1} C(a_1,b_1) \\
&= \lim_{a \downarrow 0} \frac{a}{a_1 b_1} C(a_1,b_1) \\
&= 0.
\end{aligned}$$

The derivations of $\lambda_U, \lambda_{U,L}, \lambda_{L,U}$ are a bit more involved and can be found in Appendix A. □

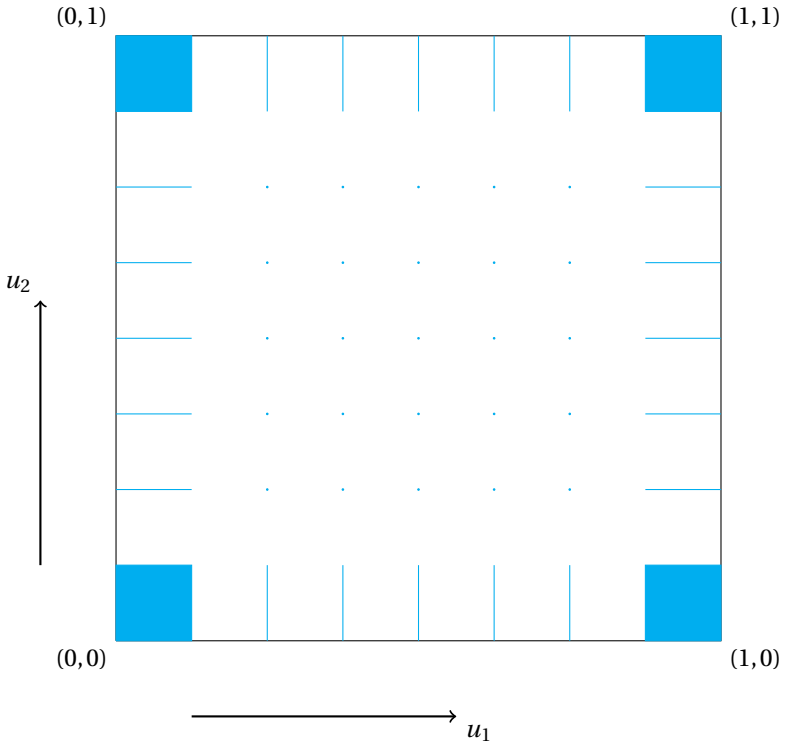
Thus, the TDCs equal 0, independent of the underlying distribution. This is undesirable when we wish to fit the copula to data where tail dependence could occur. In the next section we will change the copula, such that tail dependence can be incorporated.

4.9.1. Copula construction with tail dependence

To obtain tail dependence in a copula we will adjust the copula for values close to $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ in such a way that our copula is still a copula and tail dependence could occur. We do this via a subcopula of which the domain is given by $D = D_1 \times D_2$, where $D_1 = D_2 = \left[0, \frac{1}{m}\right] \cup \frac{2}{m} \cup \cdots \cup \frac{m-2}{m} \cup \left[\frac{m-1}{m}, 1\right]$. An example of this domain is given by Figure 4.7, where $m = 8$, corresponding to m pairs of observations. Note that we are typically dealing with significantly larger sample sizes (e.g. in the order of 100 or 1000).

Figure 4.7: **Domain of subcopula**

This figure shows the domain of the subcopula defined by equation (4.34).



Durante et al. (2017) use a similar approach, but define the copula on the squares near $(0,0)$ and $(1,1)$ only and not on the other squares, grid points and lines. They give an upper and lower bound for the extension of this copula, but do not show how we can use information we obtain from non-tail observations in our copula. Using the methodology of Durante et al. (2017) we can therefore approximate the copula on the squares, but they do not propose a method to approximate the copula outside the squares, which is what we will do here.

In Theorem 4.18 we define the subcopula on the domain of Figure 4.7 such that it can include tail dependence and that it approximates the copula on the grid points.

Theorem 4.18. *Suppose we have a sample $X_1,\cdots,X_m \in \mathbb{R}^2$ for X. Then let us define $C_S : \left([0,\frac{1}{m}]\cup\frac{2}{m},\cdots,\frac{m-2}{m}\cup[\frac{m-1}{m},1]\right)^2 \to [0,1]$ by the linear extension of the grid copula C_G on the blue dots, the blue lines and the squares including $(0,1)$ and $(1,0)$. Let us define it by the scaled comonotonic copula on the squares including $(0,0)$ and $(1,1)$ Then this gives us:*

$$C_S(u_1,u_2) = \begin{cases} m\min\{u_1,u_2\}C_G\left(\frac{1}{m},\frac{1}{m}\right), & \text{for } (u_1,u_2)\in\left[0,\frac{1}{m}\right]^2, \\ u_1+u_2-1+\min\{1-u_1,1-u_2\}\left(2-m+mC_G\left(\frac{m-1}{m},\frac{m-1}{m}\right)\right), & \text{for } (u_1,u_2)\in\left[\frac{m-1}{m},1\right]^2. \\ C_{G,L}(u_1,u_2), & \text{else.} \end{cases} \tag{4.34}$$

Then this is a subcopula.

Proof. The proof can be found in Appendix A, where this theorem is a special case of Theorem 4.21 with $a=b=1$. □

Theorem 4.19. *The linear extension of the subcopula defined by equation (4.34) based on m observations is a copula with tail dependence coefficients $\lambda_L = mC_G\left(\frac{1}{m},\frac{1}{m}\right)$, $\lambda_U = 2-m+mC\left(\frac{m-1}{m},\frac{m-1}{m}\right)$, $\lambda_{L,U} = 1-mC\left(\frac{1}{m},\frac{m-1}{m}\right)$ and $\lambda_{L,U} = 1-mC\left(\frac{m-1}{m},\frac{1}{m}\right)$.*

Proof. Any extension of a subcopula defined by equation (4.32) is a subcopula and we show the equality for

λ_L, λ_U and $\lambda_{L,U}$ (the derivation of $\lambda_{U,L}$ is similar to the derivation of $\lambda_{L,U}$):

$$\begin{aligned}
\lambda_L &= \lim_{a\downarrow 0} \frac{C(a,a)}{a} = \lim_{a\downarrow 0} \left(mC_G\left(\frac{1}{m}, \frac{1}{m}\right) \frac{\min\{a,a\}}{a} \right) = mC_G\left(\frac{1}{m}, \frac{1}{m}\right), \\
\lambda_U &= \lim_{a\uparrow 1} \frac{1-2a+C(a,a)}{1-a} = \lim_{a\uparrow 1} \frac{\min\{1-a,1-a\}\left(2-m+mC_G\left(\frac{m-1}{m}, \frac{m-1}{m}\right)\right)}{1-a} = 2-m+mC_G\left(\frac{m-1}{m}, \frac{m-1}{m}\right), \\
\lambda_{L,U} &= 1-\lim_{a\downarrow 0} \frac{C(a,1-a)}{a} = 1-\lim_{a\downarrow 0} \frac{a-\min\{a,a\}\left(1-mC_G\left(\frac{1}{m}, \frac{m-1}{m}\right)\right)}{a} = 1-mC_G\left(\frac{1}{m}, \frac{m-1}{m}\right).
\end{aligned} \tag{4.35}$$

□

Recall that, using equation (4.33), we can use the following for the linear extension of the subcopula in equation (4.34).

$$\begin{aligned}
\lambda_L &= \sum_{b=1}^{m} \mathbb{1}\left(R_1^b \leq 1, R_2^b \leq 1\right) = \hat{\lambda}_L^{(1)}(k=1), \\
\lambda_U &= 2-m+\sum_{b=1}^{m} \mathbb{1}\left(R_1^b \leq m-1, R_2^b \leq m-1\right) = \sum_{b=1}^{m} \mathbb{1}\left(R_1^b > m-1, R_2^b > m-1\right) = \hat{\lambda}_U^{(1)}(k=1), \\
\lambda_{L,U} &= 1-\sum_{b=1}^{m} \mathbb{1}\left(R_1^b \leq m-1, R_2^b \leq 1\right) = \sum_{i=1}^{n} \mathbb{1}\{R_1^i \leq 1, R_2^i > n-1\} = \hat{\lambda}_{L,U}^{(1)}(k=1), \\
\lambda_{U,L} &= 1-\sum_{b=1}^{m} \mathbb{1}\left(R_1^b \leq 1, R_2^b \leq m-1\right) = \sum_{i=1}^{n} \mathbb{1}\{R_1^i \leq n-1, R_2^i > 1\} = \hat{\lambda}_{U,L}^{(1)}(k=1),
\end{aligned}$$

where $k=1$ means the estimation is done for a threshold of 1. This estimate has a low bias, but a high standard error and we can only obtain the values 0 or 1, which is inconvenient. Another undesired effect is that we are overfitting the data, since we take every grid point into account. The first problem will be solved in section 4.9.2. The latter problem can simply be solved by considering a subgrid, which leads to Proposition 4.20.

Proposition 4.20. *Let $C_s : D_1 \times D_2 \to [0,1]$ be a 2-dimensional subcopula then for $D_1^* \subseteq D_1$ and $D_2^* \subseteq D_2$ both containing 0 and 1 we have that C_s^* defined by $C_s^* = C_s$ on $D_1^* \times D_2^*$ is a subcopula.*

Proof. This follows directly from the definition of a subcopula, since $D_1^* \subseteq D_1 \subseteq [0,1]$ and $D_2^* \subseteq D_1 \subseteq [0,1]$ and both contain 0 and 1, thus (1) must hold. Also (2), (3), (4) hold for $D_1^* \times D_2^*$ since these (in)equalities hold on a set containing $D_1^* \times D_2^*$, namely $D_1 \times D_2$. □

According to Proposition 4.20 we can safely define the subcopula on any subgrid and extend this copula linearly. This gives us freedom in choosing how many grid points we will take into account. To keep the tail dependence in the copula we will still use the different expressions for the blue squares in Figure 4.7. We can also make the squares larger when we consider different subgrids.

4.9.2. Tail dependence estimation and subcopula

To get a lower variance in the estimate of the tail dependence, we can take more grid points into account for the four corners, which corresponds to making the blue squares in Figure 4.7 larger. This is equivalent to using a larger threshold in the estimation of tail dependence. We will use this fact to generalize our subcopula.

Suppose we consider a subgrid of m times m points for the copula for which we have $d \cdot m$ data points instead of m, then the lower tail dependence coefficient will be:

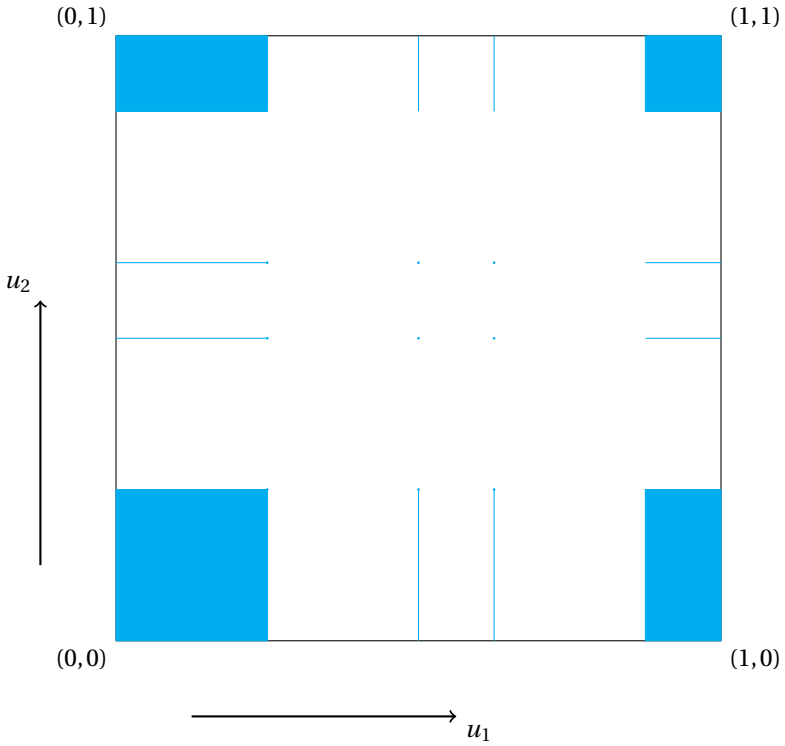
$$\lambda_L = mC_G\left(\frac{d}{dm}, \frac{d}{dm}\right) = m \cdot \frac{1}{dm} \sum_{b=1}^{dm} \mathbb{1}\left(R_1^b \leq d, R_2^b \leq d\right) = \frac{1}{d} \sum_{b=1}^{dm} \mathbb{1}\left(R_1^b \leq d, R_2^b \leq d\right) = \hat{\lambda}_L(k=d),$$

which is the estimator for the tail dependence coefficient for threshold d. Now we can get the values $\lambda_L = \{0, \frac{1}{d}, \cdots, 1\}$, which raises the bias, but significantly lowers the variance and gives more possible values of λ_L.

For the estimates of the upper TDC we get a similar result, making this an appropriate method. It also gives us freedom in the tail dependence estimation. Using this approach we need to choose the same threshold for upper and lower TDC. This means that we cannot choose the optimal threshold for both estimators.

Figure 4.8: **Domain of subcopula**

This figure shows an example of a domain of the subcopula defined by equation (A.1).



We can find a way around this problem and use a different thresholds for the estimation of the coefficients of upper and lower tail dependence. For this subcopula we define the copula on two squares of different sizes in the neighbourhood of $(0,0)$ and $(1,1)$. Due to the difference in size we need to define it on rectangles in the neighbourhood of $(0,1)$ and $(1,0)$ such that it still is a subcopula. Besides, we may freely choose a subgrid, so the grid points do not need to be equally spaced as they were in equation (4.34). An example of a domain is given by Figure 4.8. Still assuming we have m data points and taking thresholds a and b in the estimation of the coefficients of lower and upper tail dependence, we define the subcopula in Theorem 4.21.

Theorem 4.21. *Let* $D_1, D_2 \subseteq \left\{\frac{a+1}{m}, \frac{a+2}{m}, \cdots, \frac{m-b-1}{m}\right\}$ *and* $C_S : \left(\left[0, \frac{a}{m}\right] \cup D_1 \cup \left[\frac{m-b}{m}, 1\right]\right) \times \left(\left[0, \frac{a}{m}\right] \cup D_2 \cup \left[\frac{m-b}{m}, 1\right]\right) \to [0,1]$ *be defined by the linear extension of the grid copula* C_G *on the blue dots, the blue lines and the rectangles including* $(0,1)$ *and* $(1,0)$. *Let us define it by the scaled comonotonic copula on the squares including* $(0,0)$ *and* $(1,1)$. *Then this gives us:*

$$C_S(u_1,u_2) = \begin{cases} \frac{m}{a}\min\{u_1,u_2\} C_G\left(\frac{a}{m},\frac{a}{m}\right), & \text{for } (u_1,u_2) \in \left[0,\frac{a}{m}\right]^2, \\ u_1+u_2-1+\min\{(1-u_1),(1-u_2)\}\left(2-\frac{m}{b}+\frac{m}{b}C_G\left(\frac{m-b}{m},\frac{m-b}{m}\right)\right), & \text{for } (u_1,u_2) \in \left[\frac{m-b}{m},1\right]^2, \\ C_{G,L}(u_1,u_2) & \text{else.} \end{cases} \tag{4.36}$$

Then this is a subcopula.

Proof. The proof can be found in Appendix A. □

Again we can extend this copula and see whether this copula has the desired properties. First we calculate the tail dependence coefficients and give those in the following proposition.

Proposition 4.22. *Let C be the linear extension of C_s defined by equation (A.1). Then C has the following tail dependence coefficients:*

$$\begin{aligned} \lambda_L &= \hat{\lambda}_L(k=a) \\ \lambda_U &= \hat{\lambda}_U(k=b) \\ \lambda_{L,U} &= 0 \\ \lambda_{U,L} &= 0 \end{aligned} \tag{4.37}$$

Proof. The derivation of $\lambda_{L,U} = \lambda_{U,L} = 0$ is the same as in Proposition 4.17. For λ_L and λ_U we can write:

$$\lambda_L = \lim_{u\downarrow 0} \frac{C(u,u)}{u} = \lim_{u\downarrow 0}\left(\frac{m}{a} C_G\left(\frac{a}{m},\frac{a}{m}\right)\frac{\min\{u,u\}}{u}\right) = \frac{m}{a} C_G\left(\frac{a}{m},\frac{a}{m}\right) = \frac{1}{a}\sum_{j=1}^{m} \mathbb{1}\left(R_1^j \le a, R_2^j \le a\right) = \hat{\lambda}_L(k=a)$$

$$\begin{aligned} \lambda_U &= \lim_{a\uparrow 1} \frac{1-2a+C(a,a)}{1-a} \\ &= \lim_{u\uparrow 1} \frac{\min\{1-u,1-u\}\left(2-\frac{m}{b}+\frac{m}{b}C_G\left(\frac{m-b}{m},\frac{m-b}{m}\right)\right)}{1-u} \\ &= 2-\frac{m}{b}+\frac{m}{b}C_G\left(\frac{m-b}{m},\frac{m-b}{m}\right) \\ &= 2-\frac{m}{b}+\frac{1}{b}\sum_{j=1}^{m}\mathbb{1}\left(R_1^j \le m-b, R_2^j \le m-b\right) \\ &= 2-\frac{m}{b}+\frac{1}{b}\sum_{j=1}^{m} 1-\mathbb{1}\left(R_1^j > m-b \text{ or } R_2^j > m-b\right) \\ &= 2-\frac{1}{b}\sum_{j=1}^{m}\mathbb{1}\left(R_1^j > m-b \text{ or } R_2^j > m-b\right) \\ &= 2-\frac{1}{b}\left(2b-\sum_{j=1}^{m}\mathbb{1}\left(R_1^j > m-bR_2^j > m-b\right)\right) \\ &= \frac{1}{b}\sum_{j=1}^{m}\mathbb{1}\left(R_1^j > m-b, R_2^j > m-b\right) = \hat{\lambda}_U(k=b). \end{aligned}$$

□

So the TDCs of linear extension of the copula defined by equation (A.1) are equal to the lower and upper tail dependence estimators for thresholds a and b, respectively. We can therefore first perform the estimation of the TDCs where we have to choose thresholds a and b. For the subcopula defined by equation (A.1) we use the same a and b, such that the created copula has the tail dependence coefficients of the estimates.

When $\lambda_{L,U} > \lambda_U, \lambda_L$ or $\lambda_{U,L} > \lambda_U, \lambda_L$ holds, the lower-upper or upper-lower tail dependence is the strongest. It then makes more sense to base the copula estimate on these two coefficients, so that we capture this extreme dependence with our copula and set $\lambda_U = 0$ and $\lambda_L = 0$. We could then apply the copula estimation to X and $-Y$ instead of X and Y. Using equation (4.10) we can easily calculate the corresponding copula for X and Y. When all coefficients are 0 it does not make sense to use equation (A.1) and we can simply use the linear extension of the grid copula $C_{G,L}$, instead of the linear extension of C_S. The copula therefore differs for different tail dependence coefficients, which leads to the copula estimation procedure given in section 4.9.3.

4.9.3. Final copula estimation procedure

Suppose we have data $\{X_1, Y_1\}, \cdots \{X_n, Y_n\}$ available on which we wish to estimate the copula. Then we distinguish the following cases in the estimation of the copula:

(1) All TDC estimates are 0;
(2) One of the TDC estimates is non-zero;
(3a) More than one of TDC estimates are non-zero and $\max\{\lambda_L, \lambda_U\} \ge \max\{\lambda_{L,U}, \lambda_{U,L}\}$;
(3b) More than one of TDC estimates are non-zero and $\max\{\lambda_L, \lambda_U\} < \max\{\lambda_{L,U}, \lambda_{U,L}\}$.

Under (1), we can simply use the linear extension of the grid copula $C_{G,L}$.

Under (2), we can fit a copula with lower tail dependence only for the copula of (X,Y), $(X,-Y)$, $(-X,Y)$ or $(-X,-Y)$ dependent on which of the coefficients is non-zero. Lower tail dependence can be obtained by using the linear extension of the subcopula defined by equation (A.1) with $a \neq 0$ and $b = 0$. After this we can calculate the corresponding copula for (X,Y) using Proposition 4.4.

Under (3a), we can fit a copula for (X,Y) by extending the subcopula of equation (A.1).

Under (3b), we can fit a copula for $(X,-Y)$ by extending the subcopula of equation (A.1). After this we can calculate the corresponding copula for (X,Y) using Proposition 4.4.

Under (2), (3a) or (3b), we need to determine the thresholds a and/or b. These values are determined by the threshold selection method in section 4.7. We can use values of a and b such that the tail dependence coefficients implied by the choices of a and b are closest to the estimated values from the regression. Here, we should only consider small values, for example, $a, b \leq \sqrt{n}$ for n the sample size. Then a and b are given by equation (4.38).

$$\begin{aligned} a &= \underset{a\in\{1,\sqrt{n}\}}{\arg\min} |\hat{\lambda}_L^{(1)}(a) - \hat{\lambda}_L|, \\ b &= \underset{b\in\{1,\sqrt{n}\}}{\arg\min} |\hat{\lambda}_U^{(1)}(b) - \hat{\lambda}_U|, \end{aligned} \tag{4.38}$$

where $\hat{\lambda}_L$ and $\hat{\lambda}_U$ are given by the best performing regression estimators for the TDCs in our simulation study. We will find these in section 4.11.3.

4.9.4. Extension of subcopula in higher dimensional cases

The copula extension of our subcopula is now only defined for the 2-dimensional setting. Since the VAR(1) model that Ortec Finance uses to describe factors dynamics is typically m-dimensional for $m \approx 10$, we can only cover dependence between sets of 2 errors. Since we do not know a priori the dependence between the m errors, we wish to extend our 2-dimensional copula to the m-dimensional setting. To do this we can use vine copulas. A vine copula is obtained by decomposing the density function of a multi-dimensional variable into a product of conditional density functions that can all be described by a 2-dimensional copula, see Aas et al. (2009). Equation (4.39) gives an example of how we can decompose the 3-dimensional density function of random variables x_1, x_2 and x_3.

$$f(x_1,x_2,x_3) = f_1(x_1) f_2(x_2) f_3(x_3) c_{12}\left(F_1(x_1), F_2(x_2)\right) c_{23}\left(F_2(x_2), F_3(x_3)\right) c_{13|2}\left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right). \tag{4.39}$$

There are three different possible decompositions, we can namely permute x_1, x_2 and x_3 in equation (4.39). These different decompositions correspond to different vines.

An important component in the vine copula approach is the assumption of conditional independence between two variables. Assuming that x_1 and x_3 are independent given x_2 gives us $c_{13|2}\left(F_{1|2}(x_1|x_2), F_{3|2}(x_3|x_2)\right) = 1$ and therefore equation (4.39) simplifies to equation (4.40).

$$f(x_1,x_2,x_3) = f_1(x_1) f_2(x_2) f_3(x_3) c_{12}\left(F_1(x_1), F_2(x_2)\right) c_{23}\left(F_2(x_2), F_3(x_3)\right). \tag{4.40}$$

From this representation we see that we only have to estimate the copulas that describe the dependence between the pair x_1 and x_2 and the pair x_2 and x_3. Algorithms for sampling from vine copulas are given in Aas et al. (2009).

Joe et al. (2010) discusses the link between vine copulas and tail dependence, but how the tail dependence depends on the vine structure is not discussed. It is clear that different vine structures can lead to different dependence structures and this could therefore also lead to different TDCs. We should choose the vine structure such that the mistake induced by the constant conditional distribution assumption is minimal. For the three dimensional we can consider the three possible vine structures, but as the dimension of the problem increases so does the amount of possible vine structures. In five dimensions we are already dealing with 240 possible vine structures, see Aas et al. (2009), making this vine structure selection difficult for large dimensions.

4.10. Sampling from a copula

When we have determined the copula and the underlying distributions of a set of variables $X_1, \cdots, X_n$, we will need to find a method for sampling from this copula. For an arbitrary copula this can be done using the conditional distribution method, which uses the conditional copula.

Theorem 4.23. *Let C be an n-dimensional copula, strictly increasing on $(0,1)^n$ in both variables, continuously differentiable with $F_{x_1}, \cdots, F_{x_n}$ as distribution functions. Then, for $U_1, \cdots, U_n \sim U[0,1]$ independent we write $c_{u_1,\cdots u_{j-1}}(u_j) = \frac{\partial^{j-1}}{\partial u_1 \cdots \partial u_{j-1}} C(u_1, \cdots, u_n)$ and let $X_1, \cdots, X_n$ be sampled as follows:*

$$\begin{aligned} X_1 &= F_{x_1}^{-1}(U_1), \\ X_2 &= F_{x_2}^{-1}(c_{U_1}^{-1}(U_2)), \\ &\vdots \\ X_n &= F_{x_n}^{-1}(c_{U_1,\cdots,U_{n-1}}^{-1}(U_n)). \end{aligned} \tag{4.41}$$

Then we have that the dependence between $X_1, \cdots, X_n$ is described by copula C and X_i has distribution function F_{X_i} for $i = 1, 2, \cdots, n$,

Apart from a little adjustment, this theorem is the conditional distribution method explained in Schmitz (2003). They also prove that the theorem holds.

Suppose we wish to sample from the linear copula extension of the subcopula defined by (A.1). To be able to sample from the copula estimate, we need to be able to compute the inverse of c_{u_1}. This can be a problem, since this function c_{u_1} is not necessarily bijective. For an arbitrary copula it can even be the case that this function is neither injective nor surjective. Example 4.3 shows this problem.

Example 4.3. *Let C be the comonotonic copula, recall that: $C(u_1, u_2) = \min\{u_1, u_2\}$.*
Then for any given u_1 we have that $c_{u_1}(u_2)$ is neither injective nor surjective and so we cannot use the sampling method in Theorem 4.23.

Proof.

$$c_{u_1}(u_2) = \frac{\partial}{\partial u_1} C(u_1, u_2) = \frac{\partial}{\partial u_1} \min\{u_1, u_2\} = \begin{cases} 0 & \text{if } u_1 > u_2 \\ 1 & \text{else } u_1 < u_2 \end{cases}$$

Therefore, the values between 0 and 1 are not attained, meaning that C is not surjective. Besides, 0 and 1 are not uniquely attained, meaning that C is not injective. Note that the function here is not defined for $u_1 = u_2$. □

To overcome this problem, we define a generalized inverse, such that we can sample from the generalized inverse. It is defined in Definition 4.18.

Definition 4.18. *Let h be a non-decreasing function. Then we define the generalized inverse of h by:*

$$h^{\{-1\}}(x) = \inf\{u : h(u) \geq x\}, \tag{4.42}$$

which is also non-decreasing.

Using this generalized inverse we can sample from distribution functions that are not bijective. This is given in Theorem 4.24.

Theorem 4.24. *Let C be the extended subcopula of equation (A.1), with $F_{x_1}, \cdots, F_{x_n}$ as distribution functions. Then for $U_1, \cdots, U_n \sim U[0,1]$ independent, we write $c_{u_1,\cdots u_{j-1}}(u_j) = \frac{\partial^{j-1}}{\partial u_1 \cdots \partial u_{j-1}} C(u_1, \cdots, u_n)$ and let $X_1, \cdots, X_n$ be sampled as follows:*

$$\begin{aligned} X_1 &= F_{x_1}^{\{-1\}}(U_1), \\ X_2 &= F_{x_2}^{\{-1\}}(c_{U_1}^{-1}(U_2)), \\ &\vdots \\ X_n &= F_{x_n}^{\{-1\}}(c_{U_1,\cdots,U_{n-1}}^{-1}(U_n)). \end{aligned} \tag{4.43}$$

Then we have that the dependence between $X_1, \cdots, X_n$ is described by copula C and X_i has distribution function F_{X_i} for $i = 1, 2, \cdots, n$,

This is a generalized version of Theorem 4.23 and the proof can be found in Schmitz (2003).

In Example 4.3 we have seen that we could not apply Theorem 4.23 to generate random numbers from the comonotonic copula, but we will now show that we can use Theorem 4.24 for this. Let us apply this theorem to the example to see whether this sampling method works for this special case. First simply compute $c_{U_1}^{\{-1\}}(U_2)$, which gives us:

$$c_{U_1}^{\{-1\}}(U_2) = \inf\{u : c_{U_1}(u) \geq U_2\} = \inf\{u : \mathbb{1}\{U_1 < u\} \geq U_2\} = \inf\{u : u > U_1\} = U_1.$$

This means that $X_1 = F_{x_1}^{-1}(U_1)$ and $X_2 = F_{x_2}^{-1}(U_1)$, indicating that X_1 and X_2 are fully correlated, which is indeed the case for the comonotonic copula.

To find the generalized inverse for arbitrary functions we use the bisection method, see Stewart (2011). The Newton Raphson method cannot be used here because the derivative of c_{U_1} can equal zero, meaning that the method will not converge to the solution. In section 4.11.4 we sample from the copula estimate using this method. The results of the methods introduced in this chapter are given in the next section.

4.11. Results of estimations

4.11.1. Simulation study

In this section we will show the results for the tail dependence test, the tail dependence coefficients and the copula estimation. For the tail dependence test we use the Neyman-Pearson test for different α and show the results in section 4.11.2. In section 4.11.3 we give the results of the final TDC estimator, where we test the performance of the two regression methods on the two estimators $\lambda^{(1)}$ and $\lambda^{(2)}$. Recall that in both sections we do a WLS regression with the thresholds as weigths and an OLS regression. Besides, we will compare the methods with the Plateau-finding method. Lastly, in section 4.11.4 we will estimate the copula using the subcopula extension.

4.11.2. Test for tail dependence

In this section we will show the results for the tests for tail dependence based on the Clayton, Gumbel and Gaussian copulas. We know there is tail dependence for the Clayton and Gumbel copulas and tail independence for the Gaussian copula. In Tables 4.2, 4.3 and 4.4 we show the results for these three distributions for sample sizes 10000. In Appendix D we will give the results for smaller-sized data sets (300 and 1000).

Table 4.2: **Percentage of rejections of tail dependence for data from Clayton copula.**

This table shows the percentage of rejections of H_0 when the underlying is a Clayton copula. This percentage should approximate the significance level of the test. This is based on 1000 samples with sample size 10000 for $\lambda = 0.1, \cdots, 0.9$.

α ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	25.0%	6.5%	2.6%	1.6%	0.9%	0.8%	0.9%	1.0%	0.9%
3%	40.8%	14.1%	7.2%	3.9%	3.0%	2.8%	4.0%	2.6%	2.4%
5%	49.1%	21.0%	10.9%	6.9%	5.7%	3.8%	6.8%	5.5%	3.9%
10%	63.6%	35.8%	21.2%	15.2%	10.5%	8.9%	12.3%	10.4%	8.7%
20%	78.0%	51.5%	34.5%	27.9%	18.6%	18.8%	22.1%	23.2%	19.2%

Table 4.3: **Percentage of rejections of tail dependence for data from Gumbel copula.**

This table shows the percentage of rejections of H_0 when the underlying is a Gumbel copula. This percentage should approximate the significance level of the test. This is based on 1000 samples with sample size 10000 for $\lambda = 0.1, \cdots, 0.9$.

α ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1%	32.8%	6.3%	1.8%	1.1%	0.7%	1.4%	0.9%	1.0%	1.6%
3%	49.4%	12.2%	5.0%	3.6%	3.4%	4.2%	2.4%	2.7%	3.4%
5%	57.4%	17.3%	7.2%	5.9%	5.9%	6.2%	4.2%	5.1%	5.5%
10%	71.1%	27.7%	14.1%	11.6%	11.1%	11.2%	10.2%	9.8%	9.9%
20%	83.2%	45.0%	27.2%	23.1%	22.9%	22.2%	19.8%	18.9%	18.3%

In Tables 4.2 and 4.3 we would expect the percentage of rejections to be close to α. This is indeed the case for the larger tail dependence coefficients $\lambda \geq 0.5$. For the smaller tail dependence coefficients, $\lambda \leq 0.2$ we are not at all close to this expectation. This may be due to the slower convergence towards the asymptotic result of equation (4.31), which is logical since the tail dependence is smaller in these cases.

We choose to work with $\alpha = 5\%$ for the rest of the thesis, which indicates that asymptotically we can expect to reject tail dependence in 5% of the tail dependent cases. When our sample of 10000 data points has TDC $\lambda = 0.1$ we will reject H_0 with probability 49% and 57% for the underlying Clayton and Gumbel copulas, respectively. The significance level for finite samples is therefore larger than $\alpha = 5\%$.

The Gaussian copula has no tail dependence unless $\rho = 1$ and we therefore know that H_1 holds. We can therefore approximate the power $p = \mathbb{P}(C|H_1)$ by the amount of rejections under H_1 over the amount of calculations.

Table 4.4: **Percentage of rejections of tail dependence for data from Gaussian copula.**

This table shows the percentage of rejections of H_0 when the underlying is a Gaussian copula. This percentage should approximate the significance level of the test. This is based on 1000 samples with sample size 10000 for $\rho = 0, 0.1, \cdots, 1.0$.

α ρ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1%	100%	99.9%	99.9%	99.1%	93.0%	80.8%	57.6%	35.2%	15.6%	8.0%	0.8%
3%	100%	100%	100%	99.8%	97.6%	91.0%	75.3%	52.7%	29.8%	15.0%	2.5%
5%	100%	100%	100%	99.8%	99.2%	94.6%	82.6%	64.0%	37.9%	21.5%	4.4%
10%	100%	100%	100%	100%	99.9%	98.0%	90.5%	77.4%	53.5%	32.9%	9.6%
20%	100%	100%	100%	100%	99.9%	99.3%	96.0%	89.3%	71.4%	49.0%	19.7%

4.11.3. Estimation of tail dependence coefficients

In the case that our test from the previous section does not reject tail dependence, we can estimate the tail dependence coefficient. This is done for the same copulas as in the previous section. For each copula and sample size the result can be found in Appendix D. We summarize the result, average of all mean squared errors, in Table 4.5. For the value of m, the number of estimates taken into account, we chose $m = D\sqrt{n}$, where n is the sample size and we considerd various D's. Based on some test simulations we choose D to be equal to 10.

Table 4.5: **Average MSE of regressions for coeffient of tail dependence estimation on different copulas.**

This table shows the square root of the mean squared errors ($\sqrt{\text{MSE}}$) for the WLS and OLS regression of the two tail dependence estimators. For comparison we also give these results for the plateau-finding method. This is based on 2000 simulations (1000 simulations for sample size 10000) of random variables from the Clayton and Gumbel copulas with tail dependence coefficients uniformly distributed between 0 and 1.

Estimator	Regression/threshold type	300	1000	10000
$\hat{\lambda}^{(1)}$	WLS regression	† 0.060	† 0.050	† 0.029
	OLS regression	0.065	0.052	0.031
$\hat{\lambda}^{(2)}$	WLS regression	0.063	0.051	0.030
	OLS regression	0.066	0.052	0.031
$\hat{\lambda}^{(1)}$	Plateau-finding/Constant threshold	0.094	0.076	0.052
$\hat{\lambda}^{(2)}$	Plateau-finding/Constant threshold	0.086	0.075	0.049

† Denotes the best estimator for this sample size.

Note that for the size of 10000 the Plateau-finding method does not work properly, since it sometimes does not find Plateaus and sets $\hat{\lambda} := 0$ or $\hat{\lambda} := 1$ even though λ is not close to 0 or 1. The numbers in the table represent the estimates we obtain from always using $k = 0.05n = 500$ as a threshold.

In Table 4.5 we see that the WLS regression on $\hat{\lambda}^{(1)}$ performs best for all sample sizes, but the differences in MSE with the other regression methods are not that large. Also it is better than the Plateau-finding method from the literature. We will use this estimator for the rest of this thesis independent of the sample size. As expected, we also see a clear decay in the MSE when the sample size increases.

4.11.4. Estimation of copula

When we wish to estimate the copula on data we first test for tail dependence using the Neyman-Pearson test and afterwards we determine the TDCs. Next we can use the techniques in section 4.9.3 to obtain a mathematical expression for the copula. From this copula we can draw scenarios using the sampling method from 4.10. We can compare the dependence measures of these scenarios to the dependence measures of the data on which we fitted the copula. Table 4.6 shows the resulting Kendall's tau and upper TDC when using this procedure on data from underlying Gumbel copula, where we use 40 grid points for u_1 (excluding 0 and 1) and 40 grid points for u_2. Figure 4.9 shows the simulated data from this Gumbel copula and one scenario generated from the copula fitted on this data.

Table 4.6: **Comparison dependence measures of scenarios and historical data.**

This table shows important dependence measures of our 50 generated scenarios of 1000 observations and compares those to the dependence measures of the underlying sample, which are samples of 1000 observations generated from a Gumbel copula for $\lambda = 0.2, 0.4, 0.6, 0.8$.

TDC	Theoretical	0.2	0.4	0.6	0.8
	Sample	0.200	0.351	0.492	0.780
	Scenarios	0.201	0.336	0.467	0.761
	Rejection percentage of scenario	44%	38%	16%	14%
Tau	Theoretical	0.152	0.322	0.515	0.737
	Sample	0.168	0.301	0.491	0.726
	Scenarios	0.172	0.302	0.491	0.721
	MSE	0.022	0.019	0.019	0.011

In Table 4.6 we see that the TDC estimate of the scenarios are on average lower than the estimated TDCs on the sample generated from the Gumbel copula. This underestimation is due to the underestimation of the dependence caused by the linear extension. When we make the grid finer, the estimation should become better and we show this in Table 4.7.

Table 4.7: **Comparison MSEs of scenarios for different grid sizes.**

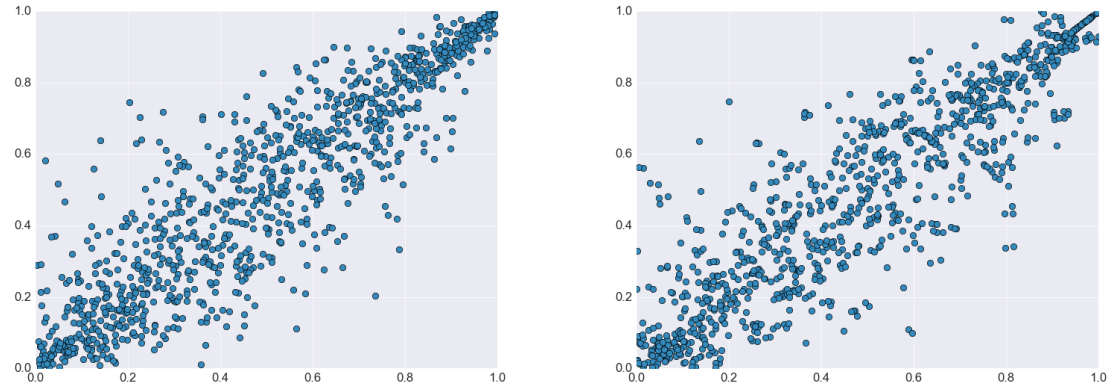
This table shows the average MSE of the scenarios $\hat{\tau}$ and λ_U for a Gumbel copula as a function of the grid size. The size of the historical and scenario data is 1000 and the grid sizes are 10, 20 and 40.

Amount of grid points	10	20	40
TDC	0.106	0.037	0.015
tau	0.029	0.018	0.018

Similar results are obtained for the dependence measures when we do the estimation on samples obtained from a Clayton copula. We show these in Tables D.7 and D.8 in the appendix. In Figure 4.9 we show one of the samples for 40 grid points and $\lambda_U = 0.8$

Figure 4.9: **Generated historical sample and scenario sample.**

This figure gives the historical sample on which we base our copula on the left and a scenario generated from this historical sample on the right. The historical sample is generated from a Gumbel copula with $\lambda_U = 0.8$



(a) Sample from Gumbel copula

(b) Sample from copula estimated on left sample

In Figure 4.9 we see that the structure of the samples looks similar on the eye and their estimated Kendall's taus are similar: $\hat{\tau} = 0.727$ versus $\hat{\tau} = 0.724$. In both samples we see clustering close to $(1,1)$, but for the scenario this clustering is on a line caused by the parametric form of the subcopula close to $(1,1)$. The estimated TDCs are nevertheless similar too: $\hat{\lambda}_U = 0.806$ versus $\hat{\lambda}_U = 0.796$. In Figure D.1 we show a scenario plot, where the underlying data is generated from a Clayton copula.

4.12. Conclusion

Many economic variables exhibit tail dependence, see for example Starica (1999). This tail dependence can be effectively captured with certain copulas that are used to describe dependence between random variables. Many parametric copula models have only one or two parameters that are fitted to the data. We cannot generally approximate the behaviour of arbitrary data by a model in which we fit only one or two parameters and we therefore come up with another approach.

In our approach we constructed a copula via an extension of a subcopula, which may include historically observed tail dependence and still accurately approximates the dependence outside the tails of the distribution. For the tail dependence estimation we introduce the WLS and OLS regression method that approximate the TDC better in MSE than the Plateau-finding method from the literature for our test cases. Since our estimation can overestimate the TDC when there is no tail dependence, we also test for tail dependence, based on the Neyman-Pearson lemma. Tail dependence is rejected more often when the TDC becomes smaller for the Clayton and Gumbel copulas and when ρ becomes smaller for Gaussian copulas. As expected, for larger sample sizes our test has the largest power, recall equation (4.29).

In conclusion, we have used a method to test for tail dependence, and came up with a method to determine the tail dependence coefficients. We used these in the semiparametric copula to approximate the historically observed dependence structure which has the estimated tail dependence coefficients. We did this for the 2-dimensional setting. We can use this copula to generate scenarios and except when tail dependence is rejected under tail dependence dependence, we estimate similar TDCs for the generated scenario as for the sample on which we base our copula. Estimations of Kendall's tau have showed that the correlation of the scenarios are similar to the correlation of the samples on which we base our copula. This suggests that this method works well for the considered data.

5

Practical implications for factor modelling

5.1. Introduction

At Ortec Finance it is important to determine the tail risks associated with all financial instruments in which clients are interested. This tail risk of a portfolio of multiple financial instruments depends on the tail risk of each individual financial instrument, for which we need to know the tail index. Also this tail risk depends on the tail dependence between the financial instruments and so we need to know the TDCs. In the model of Ortec Finance, financial and economic variables typically depend on factors and lagged factors linearly, meaning that it is important to model the factors and tail behaviour of the factors correctly.

In section 5.2 we briefly show how the economic variables depend on the factors and how the factors themselves are modelled. We have seen how copulas can be used to model the (tail) dependence between static random variables. For the factors, that are described by a VAR(1) model, we can use a copula to describe the (tail) dependence between the underlying errors in the VAR model. We will elaborate on this in section 4.4 and show what this means for the tail dependence of the factors. We should also realise that modelling the errors as independent in the VAR model already introduces some dependence structure caused by the VAR interactions. This could possibly even mean that there is tail dependence between the factors. We will elaborate on this in section 5.4. In section 5.6 we will conclude and recommend an approach for Ortec Finance.

It is interesting to see how the copula, used to describe the dependence between the errors, affects the time series. We are especially interested in the link between the tail dependence in the errors and the tail dependence in the factors. This is important for modelling the extreme behaviour of financial variables, since all financial variables depend on some of the factors through a regression, directly or indirectly. Therefore these variables will also have the tail index of at least the factor's tail index. The tail index of a financial variable will be larger when the distribution of the errrors in the regression is modelled with a larger tail index than the factor's tail index.

5.2. Ortec Finance model

Ortec Finance uses a scenario approach to predict future economic outcomes. The scenarios are generated from a dynamic factor model, which combines statistical analysis of historical data, stylized facts and expert's knowledge. In the analysis of the historical behaviour of economic variables Ortec Finance decomposes all economic variables into three frequency bands. This is done via a frequency decomposition, on which we will not elaborate here, but refer to Steehouwer (2010) for further details. We will simply assume that the data is within one of these frequency bands. In every frequency band a distinction is made between core variables and regression variables. The regression variables are typically variables for which there is not a sufficiently large sample to reliably estimate the lowest frequency component in the model. These regression variables are modelled as a function of the core variables. The core variables depend on multiple statistical factors and an error term. Here, the core variable at time t is a linear combination of these factors, lagged factors, possibly lagged values of the core variable and an error term. Ortec Finance describes the factors by a VAR model. To obtain the factors, that explain most of the behaviour of the economy, Ortec Finance uses a principal component analysis, see Jolliffe (1986). We will not elaborate on the principal component analysis here and assume the factors are given.

We summarize this dynamic factor model of Ortec Finance mathematically in Definition 5.1.

Definition 5.1 (Dynamic factor model)**.**
$y_t \in \mathbb{R}^n$ follows a dynamic factor model if it can be written as follows:

$$\begin{aligned} y_t &= \sum_{i=0}^{p} M_i f_{t-i} + \sum_{j=1}^{q} D_j y_{t-j} + \epsilon_t, \\ f_t &= A f_{t-1} + e_t, \end{aligned} \tag{5.1}$$

with $M_i \in \mathbb{R}^{n\times m}$ $\forall i$, $D_j \in \mathbb{R}^{n\times n}$ diagonal matrices $\forall j$ and $A \in \mathbb{R}^{m\times m}$ a constant matrix. $f_t \in \mathbb{R}^m$ is the vector of factors at time t, $\epsilon_t \in \mathbb{R}^n$ a vector of time independent and component wise independent and identically distributed errors and $e_t \in \mathbb{R}^m$ a vector of time independent but not necessarily independent and identically distributed error vectors at time t.

At Ortec Finance this dynamic factor model is used to describe frequency components of economic variables. The amount of factors m differs per frequency component, but typically $m \approx 10$. The amount of core variables we wish to describe with our model is much larger.

The y_t represent the financial and economic variables of interest. If we wish to model the tails of y_t correctly, we have to model the tails of f_t and ϵ_t accordingly. Ultimately, to correctly model the (tail) dependence between the components of y_t, we need to model the (tail) dependence between the factors correctly. Section 5.3 explains what to do for the former problem and section 4.4 explains how we can use copulas for the latter problem. In section 5.4 we show that heavy-tails in the errors of the VAR model can possibly induce tail dependence between the factors, but this needs mathematical back-up.

5.3. Heavy tails in VAR models

Recall we have shown the theoretical results of the tail indices of factors, that are described by a VAR model with independent errors in section 3.5. Due to the interactions between the factors caused by the matrix in the VAR model the tail of $f_{i,t}$ depends on more errors than $e_{i,t}$. As a result the tail indices of the factors equal the largest tail index of the errors. This makes the modelling of the tail difficult if we observe different tail indices for the factors. The financial time series y_t is a linear combination of its own lags, an error and the factors. We can describe the tail of the errors in the VAR model by the smallest tail index of the factors $\gamma_{\min}$, meaning that the tail index of each factor will be $\gamma_{\min}$. To get the correct tail index for y_t we can describe the error ϵ_t by a heavy-tailed distribution with the correct tail index.

If these factors were to be mutually independent, we could use Theorem 3.11 to conclude that the tail index of component y_t^i equals the largest tail index of the ϵ_t^i and the factors influencing y_t^i. The factors are however not mutually independent due to the underlying VAR model, which makes this more complicated. It is clear that the factors show some dependence implied by the VAR model. In the next section we will show some indications of this VAR model implying tail dependence.

5.4. Tail dependence between factors induced by VAR modelling

Since the factors are influenced by the errors through the VAR modelling, the factors can depend on the same errors and therefore be mutually dependent. Recall Example 3.2 where f_1 and f_2 both depended on as well ϵ_1 as ϵ_2 and we can therefore expect that f_1 and f_2 are dependent. We know that the factors are tail independent if they are modelled by independent normal errors. In this section we will investigate whether this is the case for heavy-tailed distributions as well.

To see if heavy-tailed distributions can induce tail dependence we estimate the TDC between the factors in Example 3.2 for t-distributions as well. We do this for tail indices $\gamma = 0.1, 0.2, \cdots, 0.5$. Before estimating the TDCs, we test for tail dependence and set the TDC to be $\lambda = 0$ if the test rejects tail dependence. For comparison we will also do this for a standard normal distribution (i.e. $\gamma = 0$). The results for sample sizes 300, 1000 and 10000 are given in Tables 5.1, 5.2, and 5.3, respectively. The upper and lower TDC are the same due to the symmetry of the problem. In the simulations they are very close and in the resulting tables we show the averages of the upper and lower tail results.

Table 5.1: **Tail dependence coefficient of factors described by VAR model.**

This table shows the average TDC estimates (upper and lower) of factors described by VAR model with independent underlying errors. The errors are drawn from a t-distribution with tail index $\gamma = 0.1, 0.2, \cdots, 0.5$ and standard normal distribution ($\gamma = 0$). The table shows the results for 1000 iterations of sample size 300.

Tail index	0	0.1	0.2	0.3	0.4	0.5
Percentage of rejections	40.3%	41.4%	44.9%	40.5%	35.1%	24.4%
Average non-rejected TDC estimate	0.405	0.410	0.423	0.459	0.502	0.550
Average TDC estimate	0.242	0.241	0.233	0.273	0.326	0.416

Table 5.2: **Tail dependence coefficient of factors described by VAR model.**

This table shows the average TDC estimates (upper and lower) of factors described by VAR model with independent underlying errors. The errors are drawn from a t-distribution with tail index $\gamma = 0.1, 0.2, \cdots, 0.5$ and standard normal distribution ($\gamma = 0$). The table shows the results for 1000 iterations of sample size 1000.

Tail index	0	0.1	0.2	0.3	0.4	0.5
Percentage of rejections	50.7%	51.4%	56.8%	49.2%	32.5%	19.6%
Average non-rejected TDC estimate	0.360	0.366	0.371	0.411	0.463	0.542
Average TDC estimate	0.178	0.178	0.160	0.209	0.312	0.435

Table 5.3: **Tail dependence coefficient of factors described by VAR model.**

This table shows the average TDC estimates (upper and lower) of factors described by VAR model with independent underlying errors. The errors are drawn from a a t-distribution with tail index $\gamma = 0.1, 0.2, \cdots, 0.5$ and standard normal distribution ($\gamma = 0$). The table shows the results for 200 iterations of sample size 10000.

Tail index	0	0.1	0.2	0.3	0.4	0.5
Percentage of rejections	76.0%	93.0%	93.0%	66.5%	31.0%	6.0%
Average non-rejected TDC estimate	0.268	0.269	0.277	0.322	0.418	0.547
Average TDC estimate	0.064	0.019	0.019	0.108	0.289	0.514

In the tables for the smaller sample sizes 300 and 1000 we cannot clearly reject tail dependence. For the normal distribution we know that there is no tail dependence, but we cannot always reject this from the estimates. The least rejections and largest TDC estimate is obtained for $\gamma = 0.5$. The percentage of rejections is 6.0% here which is close to the asymptotic 5%, that is expected for tail dependent data. This suggests that there is indeed tail dependence present in the data. For other tail indices we cannot conclude this. It would be interesting for further research to explore the behaviour for larger samples.

5.5. Tail dependence between factors in COPAR model

From the previous section we learned that a larger tail index for the errors in the VAR model describing the factors possibly implies a larger TDC between the factors. If we wish the tail dependence to be larger for a given tail index, we should use a copula which includes tail dependence to describe the dependence between the errors. Recall the COPAR model which is described in section 4.4. The tail dependence of the factors in a COPAR model is a function of the tail index of the errors and the tail dependence between the errors. Here, a larger tail index and a larger TDC between the errors implies a larger TDC between the factors.

Theoretically we can use an n-dimensional copula to describe the n errors in the VAR model. In practice, accurate estimation of such a copula for $n \geq 3$ is difficult, especially when we wish to have the correct tail dependence between every pair of variables. To see what happens with the tail dependence of the factors when we model the errors in the VAR model by copulas we perform some simulations for the 2-dimensional setting. We do this for the Gumbel and Clayton copulas and the results are given in Tables 5.4, 5.5 and 5.6.

Table 5.4: **Tail dependence coefficient of factors described by COPAR model.**

This table shows the TDC estimates for factors described by a COPAR model. We estimates on the factors from Example 3.2 where the errors are described by a Clayton copula and a Gumbel copula with standard normal margins. The table shows the percentage of TDC rejections and the average TDC estimate for the factors. The table shows the results for 1000 iterations of sample size 300.

TDC of errors	0	0.2	0.4	0.6	0.8
Percentage of rejections	39.3%	18.3%	9.8%	5.6%	0.6%
TDC of factors	0.404	0.551	0.625	0.713	0.825

Table 5.5: **Tail dependence coefficient of factors described by COPAR model.**

This table shows the TDC estimates for factors described by a COPAR model. We estimates on the factors from Example 3.2 where the errors are described by a Clayton copula and a Gumbel copula with standard normal margins. The table shows the percentage of TDC rejections and the average TDC estimate for the factors. The table shows the results for 500 iterations of sample size 1000.

TDC of errors	0	0.2	0.4	0.6	0.8
Percentage of rejections	50.6%	19.4%	8.6%	1.4%	0.2%
TDC of factors	0.360	0.522	0.603	0.701	0.824

Table 5.6: **Tail dependence coefficient of factors described by COPAR model.**

This table shows the TDC estimates for factors described by a COPAR model. We estimates on the factors from Example 3.2 where the errors are described by a Clayton copula and a Gumbel copula with standard normal margins. The table shows the percentage of TDC rejections and the average TDC estimate for the factors. The table shows the results for 100 iterations of sample size 10000.

TDC of errors	0	0.2	0.4	0.6	0.8
Percentage of rejections	76%	28%	12%	2%	0%
TDC of factors	0.263	0.465	0.569	0.678	0.816

As expected we see that there is a positive relation between the TDC between the errors and the TDC between the factors. Also we see that the percentage of rejections decreases as the tail dependence coefficients of the errors increase. When our TDC equals zero (in this case independent errors) we see that for the factors tail dependence is rejected 39.3%, 50.6% and 76% of the times for sample sizes 300, 1000 and 10000 respectively. From the literature we know that there is no tail dependence and we thus see that the hypothesis testing improves when our sample becomes larger but for 10000 we still do not reject 24% of the tail independent samples, but still the percentage is clearly larger than for the tail dependent samples.

We should note here that we used the Neyman-Pearson test to determine whether the data is tail dependent and the WLS regression on $\hat{\lambda}_1$ to determine the TDC estimates. Both of these methods are designed for independent samples, but the time series described by the COPAR model shows autocorrelation. The performance of the estimators on this data may therefore be worse than the performance on the errors, that do not show autocorrelation.

5.6. Conclusion and recommendations

In conclusion, to correctly model the individual tail behaviour of the core variables we recommend describing e_t in equation (5.1) by a distribution with tail index $\gamma_{\min}$ where this is the minimum of the factor's tail indices. We can also model them without heavy tail and make sure that y_t gets the right tail index through the errors ϵ_t. The constructed copula can be used to make sure that the tail dependence coefficients of the factors are correct. This modelling makes sure that we can incorporate tail dependence between the core variables by the factors, but it does not make sure that the TDCs of the core variables are the same as those observed

historically. The core variables themselves are also described by a VAR model, which causes the TDCs of the core variables to differ from the TDCs of the factors.

To have the same tail index as observed historically for the core variables, each component ϵ_t^i should have the same tail index as observed historically for y_t^i. In this way we can make sure that the individual tails of the the core variables are correct. Also we can include the tail dependence of the core variables, that is implied by the tail dependence of the factors. We can make sure that the factors have approximately the same dependence structure and tail dependence coefficients as observed historically. That this also closely approximates the tail dependence coefficients and dependence structure of the core variables cannot be concluded directly. Due to the complex interactions between the dimensions in the VAR modelling we cannot directly link the TDCs of the core variables to the TDCs of the factors.

In Table 5.6 we only reject the tail dependence hypothesis for 76% of the tail independent factors. These scenarios correspond to VAR scenarios, since the errors are described by the independence copula if $\lambda = 0$. We can then compute that the Pearson's correlation between $f_{1,t}$ and $f_{2,t}$ equals $\rho = 0.678$. From Table 4.4 we have learned that we reject tail dependence less for larger ρ-values, so it is logical that we reject only 76%. Considering larger time series should show even larger percentages, but this is computationally very time consuming. This would also give more accurate TDC estimates for the factors.

We should also note here that the copula is constructed for the 2-dimensional setting and for the extension to more dimensions the use of vine copulas seems promising, but before applying this approach it needs further research. Using the 2-dimensional copula we can therefore only describe pairs of errors and we suggest to describe the two errors that exhibit the largest tail dependence by the copula. Next, we choose the pair with the largest tail dependence from the errors that are left and so forth. Doing this leaves us with the model defined in Definition 5.2.

Definition 5.2 (Dynamic copula factor model)**.** *$y_t \in \mathbb{R}^n$ follows a dynamic copula factor model if it can be written as follows:*

$$\begin{aligned} y_t &= \sum_{i=0}^{p} M_i f_{t-i} + \sum_{j=1}^{q} D_j y_{t-j} + \epsilon_t, \\ f_t &= A f_{t-1} + e_t, \end{aligned} \tag{5.2}$$

with all variables the same as in equation (5.1), except for $e_t \in \mathbb{R}^m$ which is not necessarily componentwise independent, but still time independent. Here, the dependence between $e_{1,t}$ and $e_{2,t}$ is described by $C_{1,2}$, which is estimated using the copula estimation procedure described in section 4.9.3 applied to $e_{1,t}$ and $e_{2,t}$. The dependence between $e_{3,t}$ and $e_{4,t}$ is described by $C_{3,4}$ using the same procedure and so forth until there are no pairs left.

Using this model enables us to generate core variables with individual tails similar to what we observed historically. Besides, it enables us to incorporate tail dependence between the different core variabels via the COPAR model that describes the factors. Since we can yet only describe pairs of variables we do not capture the dependence between all the error components and therefore all the factors and core variables. It should however capture the strongest tail dependence by using copulas for the pairs which exhibit the strongest historical tail dependence. For an extension to more dimensions we can use vine copulas, which are discussed in section 4.9.4.

6

Conclusion and discussion

6.1. Conclusion

We have implemented methods to approximate the tail index and tail dependence coefficients of random variables. We also used a testing procedure to see if we are indeed dealing with tail dependence. In the context of the dynamic factor model of Ortec Finance, we can apply these methods and the testing procedure to the errors in the factor model corresponding to each frequency band, but we should take care. Under some weak conditions the tails of all factors can be described by the strongest tail index of the errors. To deal with this issue, we describe the errors by a heavy-tailed distribution with the smallest tail index of the factors $\gamma_{\min}$ as tail index. The factors will therefore have tail index of at most $\gamma_{\min}$.

To correctly model the individual tail behaviour of the core variables the errors ϵ_t in equation (5.1) should be modelled with the same tail index as the core variables themselves. To make sure that we can model tail dependence as well, we use copulas to describe the dependence between the errors e_t in equation (5.1). In this way we approximate the tail dependence between the errors, but whether this also models the factor's tail dependence similar to what we observed historically remains the question. Since the copula is suitable for the 2-dimensional setting we can only capture the dependence between pairs.

6.2. Tail index estimation

In finding a tail index estimation procedure we assumed that the extremes of the factors were of some parametric form and we considered a limited number of possible tail indices. These heavy-tailed parametric forms do not hold generally and whether the data is heavy-tailed remains the question. We use the Hill, Moment, Adjusted Hill and Regression estimators for the estimation and take an average of three of those to improve the estimate. We have only considered equally weighted averages of estimators here. Using model averaging, in which we can consider any linear combinations of these four estimators, can possibly improve the estimator.

6.2.1. Heavy-tailed underlying distribution assumption

We have implemented the tail index estimation procedure on samples from three families of distributions, namely the Pareto, Burr and Fréchet family. For arbitrary financial data it is not guaranteed that one of these is the underlying distribution and therefore the estimates for financial data could have a larger (but possibly also smaller) MSE than estimates for the simulated data. We have also considered data with tail indices $\gamma = 0.1, 0.2, \cdots, 0.5$ only and no values in between. We know that our final estimator performs well for these values of γ and it is therefore likely to perform well for values between 0.1 and 0.5.

In Haan and Ferreira (2006) it is discussed that $\gamma > 0$ is typically the case for financial time series, but we cannot be entirely sure whether all the factors show this heavy-tailed behaviour. For light-tailed data (i.e. $\gamma \leq 0$) we know that our estimator overestimates γ and should therefore not be used. To make sure that we do not use the estimator on light-tailed data we could first test whether indeed $\gamma > 0$. We propose to use the Moment estimator for this, since it is the only estimator we considered, which can also estimate $\gamma \leq 0$. If this Moment estimator gives $\gamma > 0$ we can proceed with the tail index estimation, otherwise we should conclude that the data is possibly not heavy-tailed and we do not need to model it as such.

Considering underlying distributions with $\gamma > 0.5$ for our test data and choosing the method based on this data might make the estimator more robust, but it could also mean that we lose performance (larger MSE) with respect to the currently used estimator for $0.1 \leq \gamma \leq 0.5$, which is undesirable. For further research it might be interesting to test whether we can make the estimator more robust without losing too much accuracy for $0.1 \leq \gamma \leq 0.5$. We however think it is unlikely that we have factors with $\gamma \geq 0.5$, since this would mean that the variance of the core variables that depend on these factors does not exist.

6.2.2. Model averaging

In our analysis of choosing a linear combination of estimates, we only considered unweighted averages of two, three or four estimators. We could however choose a weighted average too, which increases the degrees of freedom we have and the parameters we have to fit in the procedure. More degrees of freedom implies that we can possibly get an estimate with smaller MSE on our test data, but here we might be overfitting. Model averaging deals with this trade-off and can be applied in our tail index estimation. In the model averaging framework we can also first make an initial guess on γ and then decide on using an estimator or a linear combination, which has a small MSE for γ's close to the initial guess. This could be an interesting topic for further research.

6.3. Regression in TDC estimation

In our decision on which estimator to use we use simulated data to represent financial data. This data is simulated from two parametric families, namely Clayton and Gumbel copula. These parametric forms do not hold for arbitrary financial data, and therefore have a different rate of convergence towards the TDC. The methods we propose could therefore perform worse (or better), in MSE sense, than on data from these parametric families. For the TDC estimation, using $\hat{\lambda}^{(1)}$ and $\hat{\lambda}^{(2)}$ we used an OLS and WLS regression. In both methods we however know that there is an underlying assumption, which does not generally hold.

In both regressions we assume that the bias increases linearly with the threshold, which is not necessarily the case. To deal with this we could generalize to a polynomial regression method to see if we can improve our MSE. We did this for second degree polynomial regression, but the resulting MSEs became larger. For further research it will be interesting to see if taking more degrees into account will significantly improve the estimation.

We also know that the variance in the estimates decreases as a function of the threshold. In the OLS regression we do not account for this relation and in the WLS regression we assume this relation to be linear. Since the relation depends on the underlying, we do not know which weights to choose a priori. Also we should take into account that there is mutual dependence between the estimates. If we take these two points into account, we should be able to improve the regression. For further research we recommend using a generalized least squares regression to deal with these two problems.

6.4. Theoretical behaviour of factors with underlying copula

The theoretical behaviour of the tails in VAR models is clear when the error components are independent and identically distributed. When however the error components are dependent and described by a copula the situation becomes much more involved. We did not find a mathematical expression which links the TDCs of the copula describing the dependence between the error components to the TDCs of the factors. Based on simulation results we have found strong indications that larger tail dependence in the errors implies larger tail dependence in the factors. We have also found indications, less strong however, that a larger tail index for the errors implies stronger tail dependence in the factors.

For further research it will be interesting to see if we can link the factor TDCs to the error TDCs and tail indices and if we can mathematically prove that tail dependence between the errors implies tail dependence between the factors.

6.5. Copula estimation

In our approach we fit a copula to the historical data, which is in our case simulated, and we have some choices to make here. One of these choices is the subgrid, which can contain a small number of points, which reduces the computational effort at the cost of accuracy in describing the data. Taking more grid points into account makes the method more time consuming, but should describe the data more accurately. When we take too many grid points into account we will however overfit the data, which we wish to prevent. In this

thesis we only considered equally spaced subgrids, but we could also use unequally spaced subgrids. We could for example make the grid finer in the tails of the distribution when the tail exhibits tail dependence, such that the TDCs of our generated scenarios will resemble the TDCs of the historical data more closely.

The extension we use to extend our subcopula from the subdomain to $[0,1]^n$ is a linear extension. The linear extension has constant density on each square between four points from the subdomain. This is what we expect for independent data, but not for data with positive or negative correlation. Therefore we will underestimate this positive or negative correlation, especially when we do not take enough grid points into account, which leads to underfitting. Taking many grid points into account reduces this effect, but overfitting could then take place.

We could also use other, possibly more advanced extensions here, but we should be careful, since naively extending could result in an extension which is not a copula. In the copula estimation procedure I considered using a cubic and quintic spline instead of a linear spline between the grid points. This however gave a function that was not 2-increasing and therefore not a copula. Durante et al. (2017) determines upper and lower bounds for the extension where the copula is given on a rectangle including $(0,0)$ and a rectangle including $(1,1)$ only. Our extension is different since it extends a subcopula, which is given on the grid points as well. Using a similar approach we should be able to find upper and lower bounds for our extension as well.

For further research it will be interesting to see if we can use these extensions and possible other extensions to obtain a copula from the subcopula, which does not underestimate the correlation. Also the trade-off between the overfitting and underfitting should be taken into consideration here to make a good decision on the amount of grid points which will be used. Lastly, we can consider unequally spaced grids to better describe the tails and generalize our method even further.

6.6. Dimensionality

Dealing with large dimensionality is a challenge in copula theory. The copula we introduced accurately describes the dependence for the 2-dimensional case. We can therefore only describe dependence between pairs. To be able to deal with more dimensions we could apply vine copulas. The higher the dimension, the more difficult it is to make an accurate approximation and draw representative scenarios. It will be interesting to see with what level of accuracy we can describe (tail) dependence between random variables for higher dimensions by combining the copula designed in this thesis with vine copulas.

A

Proofs

In this part of the appendix we will give the proofs or part of the proofs of some theorems and propositions, that we have not given in the main text.

Proposition 4.17 *The linear extension of the grid subcopula, defined by equation (4.33) has the following TDCs:*

$$\lambda_L = \lambda_U = \lambda_{L,U} = \lambda_{U,L} = 0$$

.

Proof.

$$\begin{aligned}
\lambda_U &= \lim_{a\uparrow 1} \frac{1-2a+C(a,a)}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+\frac{a-a_{m-1}}{1-a_{m-1}}\frac{a-b_{m-1}}{1-b_{m-1}}+\frac{a-a_{m-1}}{1-a_{m-1}}\frac{1-a}{1-b_{m-1}}b_{m-1}+\frac{a-b_{m-1}}{1-b_{m-1}}\frac{1-a}{1-a_{m-1}}a_{m-1}+\frac{(1-a)^2}{(1-a_{m-1})(1-b_{m-1})}C(a_{m-1},b_{m-1})}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+(ma-m+1)^2+(ma-m+1)(1-a)(m-1)+(ma-m+1)(1-a)(m-1)+(1-a)^2m^2C(\frac{m-1}{m},\frac{m-1}{m})}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+(ma-m+1)^2+(ma-m+1)(1-a)(m-1)+(ma-m+1)(1-a)(m-1)}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+(ma-m+1)(m-ma-1+2a)}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+-m^2(a-1)^2+m(a-1)(2a-2)+2a-1}{1-a} \\
&= \lim_{a\uparrow 1} \frac{1-2a+(-m^2+2m)(a-1)^2+2a-1}{1-a} \\
&= (-m^2+2m)\lim_{a\uparrow 1}\frac{(1-a)^2}{1-a} = 0
\end{aligned}$$

$$\begin{aligned}
\lambda_{L,U} &= 1-\lim_{a\downarrow 0}\frac{C(a,1-a)}{a} \\
&= 1-\lim_{a\downarrow 0}\frac{a_1\frac{a-a_0}{a_1-a_0}\frac{1-a-b_{m-1}}{1-b_{m-1}}+C(a_1,b_{m-1})\frac{a-a_0}{a_1-a_0}\frac{a}{1-b_{m-1}}}{a} \\
&= 1-\lim_{a\downarrow 0}\frac{a(1-am))+C(a_1,b_{m-1})a^2m}{a} \\
&= 1-\lim_{a\downarrow 0}\{(1-ma)+C(a_1,b_{m-1})am\} \\
&= 0
\end{aligned}$$

The derivation of $\lambda_{U,L} = 0$ is equivalent due to the symmetry and the derivation of λ_L is given right after Proposition 4.17. □

Theorem 4.21 *Let* $D_1, D_2 \subseteq \left\{\frac{a+1}{m}, \frac{a+2}{m}, \cdots, \frac{m-b-1}{m}\right\}$ *and* $C_S : \left([0, \frac{a}{m}] \cup D_1 \cup \left[\frac{m-b}{m}, 1\right]\right) \times \left([0, \frac{a}{m}] \cup D_2 \cup \left[\frac{m-b}{m}, 1\right]\right) \to [0,1]$ *be defined by the linear extension of the grid copula* C_G *on the blue dots, the blue lines and the rectangles including* $(0,1)$ *and* $(1,0)$. *Let us define it by the scaled comonotonic copula on the squares including* $(0,0)$ *and* $(1,1)$. *Then this gives us:*

$$C_S(u_1,u_2) = \begin{cases} \frac{m}{a}\min\{u_1,u_2\} C_G\left(\frac{a}{m},\frac{a}{m}\right), & \text{for } (u_1,u_2) \in \left[0,\frac{a}{m}\right]^2, \\ u_1+u_2-1+\min\{(1-u_1),(1-u_2)\}\left(2-\frac{m}{b}+\frac{m}{b}C_G\left(\frac{m-b}{m},\frac{m-b}{m}\right)\right), & \text{for } (u_1,u_2) \in \left[\frac{m-b}{m},1\right]^2, \\ C_{G,L}(u_1,u_2) & \text{else.} \end{cases} \tag{A.1}$$

Then this is a subcopula.

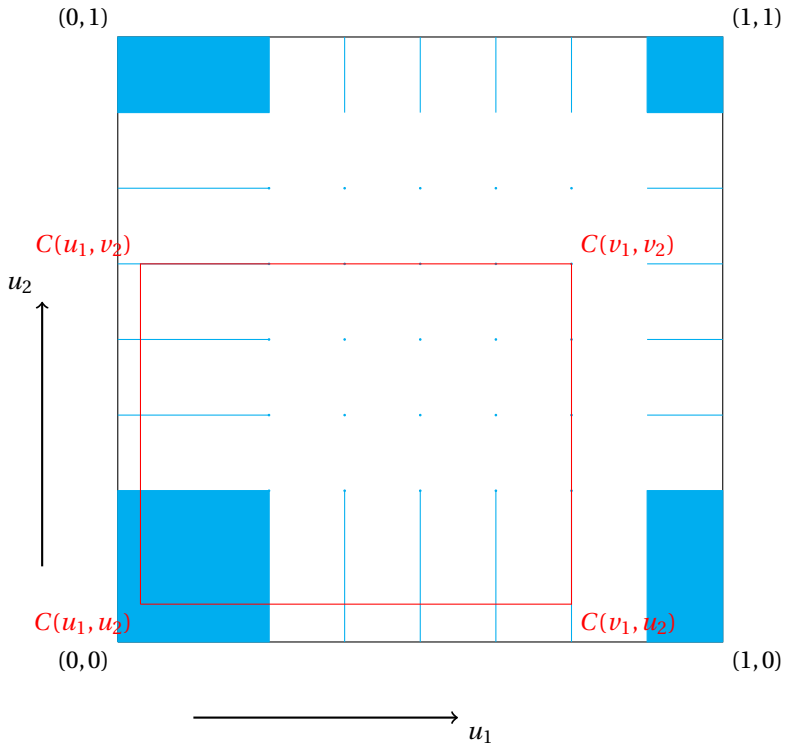
Proof. We will give the proof for the $D_1 = D_2 = \left\{\frac{a+1}{m}, \frac{a+2}{m}, \cdots, \frac{m-b-1}{m}\right\}$ case. The $D_1, D_2 \subseteq \left\{\frac{a+1}{m}, \frac{a+2}{m}, \cdots, \frac{m-b-1}{m}\right\}$ then follows from Proposition 4.20.

By construction C_S is continuous on its domain. To prove this is a copula we need to prove (1), (2), (3) and (4) of Definition 4.17 of which the first three are clearly satisfied.
So it is left to prove that $\forall (u_1,u_2)^T, (v_1,v_2)^T$, with $u_1 \le v_1$ and $u_2 \le v_2$. To do this we first define the quantity $M(u_1,u_2,v_1,v_2)$ in equation (A.2).

$$M(u_1,u_2,v_1,v_2) = C(v_1,v_2) + C(u_1,u_2) - C(u_1,v_2) - C(v_1,u_2) \tag{A.2}$$

Any M can be seen as a function of a rectangle of which the corners should be in the domain of the subcopula for which we have to show that $M(u_1,u_2,v_1,v_2) \ge 0$. We show this in Figure A.1.

Figure A.1: Domain of subcopula



Claim: For u_1, u_2, v_1, v_2 arbitrary in the domain of C_S, such that $u_1 \le v_1$ and $u_2 \le v_2$, we can write $M(u_1,u_2,v_1,v_2)$ as follows:

$$M(u_1,u_2,v_1,v_2) = \sum_{i=1}^{k} M\left(u_1^i, u_2^i, v_1^i, v_2^i\right), \tag{A.3}$$

where $u_1^k, u_2^k, v_1^k, v_2^k$ are such that $u_1^k \le v_1^k$ and $u_2^k \le v_2^k \forall k$ and (i), (ii), (iii), (iv), (v), (vi) or (vii) holds $\forall k$, where (i)-(vii) are defined below:

$$
\begin{aligned}
&\text{(i) } u_1, u_2, v_1, v_2 \notin \left[0, \frac{a}{m}\right) \cup \left(\frac{m-b}{m}, 1\right];\\
&\text{(ii) } u_1, u_2, v_1, v_2 \in \left[0, \frac{a}{m}\right];\\
&\text{(iii) } u_1, u_2, v_1, v_2 \in \left[\frac{m-b}{m}, 1\right];\\
&\text{(iv) } u_1, v_1 \in \left[0, \frac{a}{m}\right], u_2, v_2 \in \left[\frac{m-b}{m}, 1\right];\\
&\text{(v) } u_2, v_2 \in \left[0, \frac{a}{m}\right], u_1, v_1 \in \left[\frac{m-b}{m}, 1\right];\\
&\text{(vi) } u_1, v_1 \in \left[0, \frac{a}{m}\right] \cup \left[\frac{m-b}{m}, 1\right]; u_2, v_2 \in \left[\frac{a}{m}, \frac{m-b}{m}\right];\\
&\text{(vii) } u_2, v_2 \in \left[0, \frac{a}{m}\right] \cup \left[\frac{m-b}{m}, 1\right]; u_1, v_1 \in \left[\frac{a}{m}, \frac{m-b}{m}\right].
\end{aligned}
$$

When (i) holds we have:

$$
\begin{aligned}
M(u_1, u_2, v_1, v_2) &= C_S(v_1, v_2) + C_S(u_1, u_2) - C_S(u_1, v_2) - C_S(v_1, u_2)\\
&= C_{G,L}(v_1, v_2) + C_{G,L}(u_1, u_2) - C_{G,L}(u_1, v_2) - C_{G,L}(v_1, u_2)\\
&\ge 0,
\end{aligned}
$$

since $C_{G,L}$ is a subcopula.
When (ii) holds we find:

$$
M(u_1, u_2, v_1, v_2) = \frac{m}{a} C_G\left(\frac{a}{m}, \frac{a}{m}\right)(\min\{v_1, v_2\} + \min\{u_1, u_2\} - \min\{u_1, v_2\} - \min\{v_1, u_2\}) \ge 0,
$$

where the inequality follows by distinguishing the cases:

$$
\begin{aligned}
&(a)\ u_1, v_1 \le u_2, v_2;\\
&(b)\ u_1, v_1 \ge u_2, v_2;\\
&(c)\ u_2 \le u_1 \le v_1 \le v_2;\\
&(d)\ u_2 \le u_1 \le v_2 \le v_1;\\
&(e)\ u_1 \le u_2 \le v_1 \le v_2.
\end{aligned}
$$

When (iii) holds we get:

$$
M(u_1, u_2, v_1, v_2) = \left(2 - \frac{m}{b} + \frac{m}{b} C\left(\frac{m-b}{m}, \frac{m-b}{m}\right)\right)(\min\{u_1, u_2\} + \min\{v_1, v_2\} - \min\{u_1, v_2\} - \min\{v_1, u_2\}) \ge 0,
$$

where, again, the inequality follows from the cases (a),$\cdots$,(e).
When (iv) holds (the (v) case is similar due to symmetry) we find:

$$
M(u_1, u_2, v_1, v_2) = C_{G,L}(v_1, v_2) + C_{G,L}(u_1, u_2) - C_{G,L}(u_1, v_2) - C_{G,L}(v_1, u_2) \ge 0,
$$

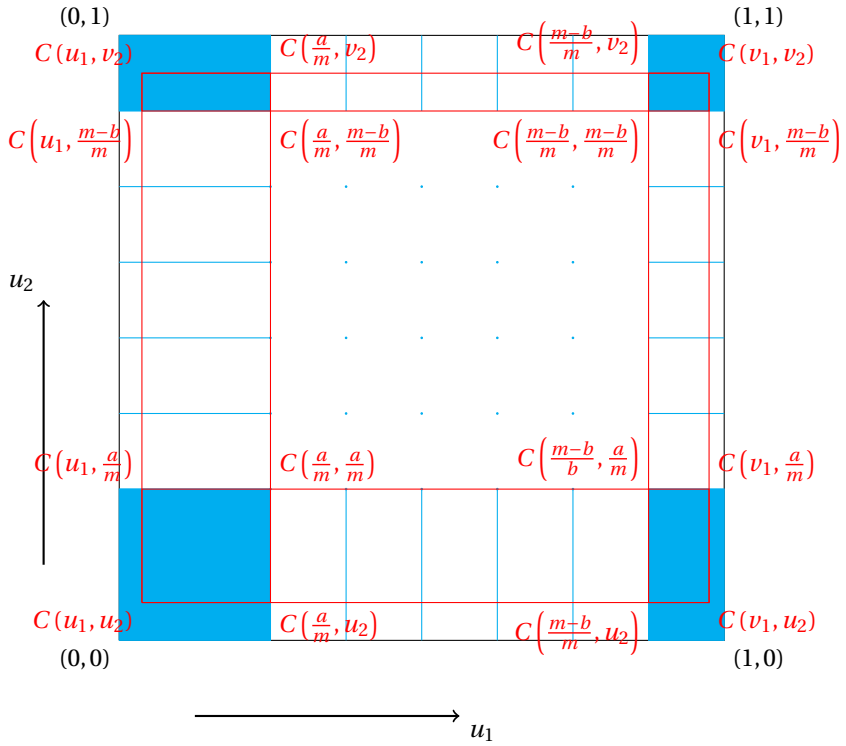
Similarly when (vi) holds (the (vii) case is similar due to symmetry) we get:

$$
M(u_1, u_2, v_1, v_2) = C_{G,L}(v_1, v_2) + C_{G,L}(u_1, u_2) - C_{G,L}(u_1, v_2) - C_{G,L}(v_1, u_2) \ge 0,
$$

Then $M\left(u_1^i, u_2^i, v_1^i, v_2^i\right) \ge 0 \forall i$, so $M(u_1, u_2, v_1, v_2) = \sum_{i=1}^k M\left(u_1^i, u_2^i, v_1^i, v_2^i\right) \ge 0$, which proves (4) of Definition 4.17. So all properties of a subcopula hold true, but we still need to prove the claim.

We can prove the claim for the most advanced case, namely the case that $u_1, u_2 \in (0, \frac{a}{m})$ and $v_1, v_2 \in \left(\frac{m-b}{m}, 1\right)$. By Figure A.2 we show that we can write the rectangle corresponding to M as a sum of 9 other rectangles.

Figure A.2: Domain of subcopula



For each of these rectangles (i),(ii)(iii),(iv) or (v) holds and so the claim follows for the most advanced case, where we need $k = 9$ for the nine rectangles. Mathematically we have that:

$$\begin{aligned} M(u_1,u_2,v_1,v_2) = {} & M\left(u_1,u_2,\frac{a}{m},\frac{a}{m}\right) + M\left(u_1,\frac{a}{m},\frac{a}{m},\frac{m-b}{m}\right) + M\left(u_1,\frac{m-b}{m},\frac{a}{m},v_2\right) + M\left(\frac{a}{m},u_2,\frac{m-b}{m},\frac{1}{m}\right) \\ & + M\left(\frac{a}{m},\frac{a}{m},\frac{m-b}{m},\frac{m-b}{m}\right) + M\left(\frac{a}{m},\frac{m-b}{m},\frac{m-b}{m},v_2\right) + M\left(\frac{m-b}{m},u_2,v_1,\frac{a}{m}\right) \\ & + M\left(\frac{m-b}{m},\frac{a}{m},v_1,\frac{m-b}{m}\right) + M\left(\frac{m-b}{m},\frac{m-b}{m},v_1,v_2\right), \end{aligned}$$

which is clearly of the same form as equation (A.3). The other cases are not considered in this proof, but follow more easily with $k < 9$, so the claim is proven. Thus, C_s is a subcopula. □

B

Extreme value theory results

In this part of the appendix we will present the tables from which we have obtained our major results on the threshold selection choice and the estimation of the second order parameter.

B.1. Linear combinations of estimators

Tables B.1, B.2 and B.3 show the average mean squared errors of the final estimators. These estimates are the average of 1000 estimates for samples from Burr, Pareto and Féchet distributions.

Table B.1: **Average mean squared errors of the estimators.**

This table shows the square root of the average mean squared error $\sqrt{\text{AMSE}}$ for the four estimators and some linear combinations for a sample size of 300.

Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill (H)	0.137	0.126	0.122	0.130	0.137	0.131
Moment (M)	0.112	0.134	0.141	0.145	0.147	0.136
Regression (R)	0.132	0.132	0.129	0.149	0.165	0.142
Adjusted Hill (A)	0.133	0.126	0.124	0.136	0.150	0.134
HM	0.119	0.119	0.119	0.129	0.139	0.125
HA	0.134	0.125	0.121	0.130	0.140	0.130
HR	0.132	0.125	0.120	0.132	0.142	0.131
MA	0.111	0.119	0.123	0.132	0.141	0.125
MR	0.112	0.123	0.126	0.140	0.149	0.131
AR	0.130	0.125	0.121	0.134	0.148	0.132
MRA	0.115	0.119	0.121	0.133	0.143	0.127
HRA	0.132	0.124	0.119	0.130	0.141	0.129
HMA	0.116	0.119	0.120	0.131	0.139	0.125
HMR	0.116	0.117	0.119	0.128	0.136	0.124
All	0.119	0.119	0.119	0.129	0.139	0.125

Table B.2: **Average mean squared errors of the estimators.**

This table shows the square root of the average mean squared error $\sqrt{\text{AMSE}}$ for the four estimators and some linear combinations for a sample size of 1000.

Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill (H)	0.118	0.104	0.089	0.088	0.099	0.099
Moment (M)	0.086	0.097	0.102	0.102	0.106	0.099
Regression (R)	0.106	0.096	0.096	0.108	0.109	0.103
Adjusted Hill (A)	0.109	0.099	0.087	0.088	0.102	0.095
HM	0.095	0.090	0.083	0.088	0.097	0.091
HA	0.113	0.101	0.086	0.087	0.098	0.098
HR	0.111	0.097	0.088	0.092	0.099	0.098
MA	0.085	0.088	0.086	0.091	0.099	0.090
MR	0.084	0.087	0.089	0.097	0.102	0.092
AR	0.106	0.096	0.087	0.093	0.100	0.097
MRA	0.089	0.088	0.085	0.091	0.098	0.090
HRA	0.110	0.097	0.086	0.089	0.097	0.096
HMA	0.091	0.088	0.084	0.090	0.097	0.090
HMR	0.093	0.090	0.083	0.087	0.097	0.090
All	0.095	0.090	0.083	0.088	0.097	0.091

Table B.3: **Average mean squared errors of the estimators.**

This table shows the square root of the average mean squared error $\sqrt{\text{AMSE}}$ for the four estimators and some linear combinations for a sample size of 10000.

Estimator	0.1	0.2	0.3	0.4	0.5	mean
Hill (H)	0.088	0.067	0.058	0.054	0.052	0.065
Moment (M)	0.058	0.056	0.062	0.056	0.057	0.058
Regression (R)	0.075	0.059	0.050	0.054	0.057	0.060
Adjusted Hill (A)	0.077	0.061	0.053	0.053	0.053	0.060
HM	0.066	0.053	0.049	0.051	0.051	0.054
HA	0.082	0.063	0.055	0.053	0.051	0.062
HR	0.081	0.062	0.053	0.052	0.052	0.061
MA	0.057	0.049	0.051	0.051	0.052	0.052
MR	0.055	0.049	0.050	0.052	0.055	0.052
AR	0.076	0.059	0.050	0.052	0.052	0.058
MRA	0.060	0.050	0.048	0.051	0.052	0.052
HRA	0.079	0.061	0.052	0.052	0.051	0.060
HMA	0.063	0.052	0.049	0.051	0.052	0.054
HMR	0.064	0.052	0.050	0.051	0.050	0.054
All	0.066	0.053	0.049	0.051	0.051	0.054

C

VAR results

This section shows the remaining results of the vector autoregressive models, that were not presented in Chapter 3.

Table C.1: **Average tail index estimates of the final estimate on a time series of** 600 **data points.**

This table shows the average of the final estimator taken over 500 tail index calculations. The calculations are done on a AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index				
ϕ	0.1	0.2	0.3	0.4	0.5
-0.9	0.136	0.178	0.191	0.265	0.229
-0.6	0.192	0.251	0.395	0.425	0.499
-0.3	0.210	0.263	0.426	0.511	0.568
0	0.223	0.269	0.412	0.527	0.610
0.3	0.217	0.258	0.397	0.517	0.537
0.6	0.163	0.196	0.311	0.459	0.501
0.9	0.094	0.146	0.172	0.276	0.309

Table C.2: **Average tail index estimates of the final estimate on a time series of** 2000 **data points.**

This table shows the average of the final estimator taken over 200 tail index calculations. The calculations are done on a AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index				
ϕ	0.1	0.2	0.3	0.4	0.5
-0.9	0.121	0.206	0.194	0.261	0.347
-0.6	0.158	0.282	0.329	0.407	0.471
-0.3	0.200	0.347	0.369	0.406	0.560
0	0.198	0.344	0.394	0.443	0.589
0.3	0.197	0.316	0.385	0.410	0.547
0.6	0.169	0.272	0.349	0.355	0.402
0.9	0.097	0.159	0.194	0.199	0.232

Table C.3: **Average tail index estimates of the final estimate on a time series of** 20000 **data points.**

This table shows the average of the final estimator taken over 200 tail index calculations. The calculations are done on a AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index				
ϕ	0.1	0.2	0.3	0.4	0.5
-0.9	0.114	0.206	0.271	0.361	0.455
-0.6	0.161	0.257	0.315	0.409	0.520
-0.3	0.187	0.275	0.349	0.439	0.543
0	0.195	0.279	0.362	0.448	0.548
0.3	0.189	0.273	0.355	0.443	0.544
0.6	0.178	0.251	0.336	0.428	0.541
0.9	0.102	0.182	0.309	0.354	0.466

Table C.4: **MSE of the final estimate on a time series of** 600 **data points.**

This table shows the square root of the mean squared error ($\sqrt{\text{MSE}}$) for the final estimate on 500 AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index					
ϕ	0.1	0.2	0.3	0.4	0.5	average
-0.9	0.065	0.041	0.162	0.209	0.300	0.182
-0.6	0.110	0.072	0.135	0.107	0.088	0.105
-0.3	0.133	0.092	0.158	0.148	0.122	0.132
0	0.145	0.100	0.153	0.161	0.152	0.144
0.3	0.135	0.088	0.134	0.149	0.106	0.125
0.6	0.084	0.073	0.121	0.137	0.108	0.107
0.9	0.051	0.102	0.172	0.183	0.234	0.162

Table C.5: **MSE of the final estimate on a time series of** 2000 **data points.**

This table shows the square root of the mean squared error ($\sqrt{\text{MSE}}$) for the final estimate on 200 AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index					
ϕ	0.1	0.2	0.3	0.4	0.5	average
-0.9	0.043	0.043	0.148	0.179	0.203	0.141
-0.6	0.072	0.095	0.075	0.075	0.080	0.080
-0.3	0.111	0.154	0.096	0.070	0.089	0.108
0	0.111	0.154	0.112	0.082	0.116	0.117
0.3	0.108	0.125	0.107	0.070	0.077	0.099
0.6	0.083	0.088	0.095	0.094	0.139	0.101
0.9	0.045	0.071	0.149	0.227	0.308	0.187

Table C.6: **MSE of the final estimate on a time series of** 20000 **data points.**

This table shows the square root of the mean squared error ($\sqrt{\text{MSE}}$) for the final estimate on 200 AR(1) time series with $\phi = -0.9, -0.6, \cdots, 0.9$.

	Tail index					
ϕ	0.1	0.2	0.3	0.4	0.5	average
-0.9	0.025	0.026	0.048	0.068	0.098	0.060
-0.6	0.066	0.062	0.031	0.031	0.047	0.049
-0.3	0.091	0.078	0.056	0.048	0.056	0.068
0	0.098	0.082	0.069	0.055	0.061	0.075
0.3	0.093	0.077	0.061	0.051	0.061	0.070
0.6	0.082	0.055	0.050	0.048	0.076	0.064
0.9	0.030	0.041	0.047	0.084	0.097	0.065

D

Copulas and tail dependence results

In this section we present some results of Chapter 4 that are not presented in the chapter. First we show the tables that summarize the results for the estimation of the coefficients of tail dependence in section D.1. In section D.2 we will give the results for the copula estimation.

D.1. Coefficients of tail dependence

In this section we give the results of the estimations of the coefficients of tail dependence on samples drawn from a Clayton and a Gumbel copula per tail dependence coefficient. For sample sizes of 300 the results are given in Tables D.1 and D.2. For sample sizes of 1000 the results are given in Tables D.3 and D.4. Lastly, for sample sizes of 10000 the results are given in Tables D.5 and D.6.

Table D.1: **Regression results for coefficient of tail dependence estimation of Clayton copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 200 simulations of 300 pairs of random variables drawn from a Clayton copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.188	0.249	0.328	0.397	0.479	0.570	0.677	0.783	0.893
		$\sqrt{\text{MSE}}$	0.106	0.083	0.068	0.069	0.071	0.067	0.056	0.044	0.023
$\hat{\lambda}^{(1)}$	OLS	mean	0.177	0.246	0.316	0.395	0.477	0.577	0.676	0.784	0.886
		$\sqrt{\text{MSE}}$	0.100	0.086	0.073	0.081	0.082	0.066	0.059	0.042	0.029
$\hat{\lambda}^{(2)}$	WLS	mean	0.204	0.263	0.339	0.407	0.488	0.579	0.686	0.792	0.901
		$\sqrt{\text{MSE}}$	0.119	0.091	0.072	0.069	0.070	0.065	0.056	0.044	0.025
$\hat{\lambda}^{(2)}$	OLS	mean	0.187	0.254	0.322	0.400	0.482	0.581	0.681	0.789	0.890
		$\sqrt{\text{MSE}}$	0.107	0.090	0.074	0.080	0.081	0.065	0.058	0.041	0.028

Table D.2: **Regression results for coefficient of tail dependence estimation of Gumbel copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 200 simulations of 300 pairs of random variables drawn from a Gumbel copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.097	0.195	0.289	0.392	0.483	0.592	0.690	0.785	0.892
		$\sqrt{\text{MSE}}$	0.057	0.064	0.065	0.059	0.063	0.053	0.045	0.042	0.022
$\hat{\lambda}^{(1)}$	OLS	mean	0.101	0.191	0.285	0.386	0.477	0.587	0.689	0.783	0.888
		$\sqrt{\text{MSE}}$	0.064	0.078	0.078	0.074	0.076	0.062	0.054	0.043	0.026
$\hat{\lambda}^{(2)}$	WLS	mean	0.118	0.211	0.301	0.400	0.489	0.597	0.693	0.787	0.894
		$\sqrt{\text{MSE}}$	0.057	0.063	0.063	0.058	0.062	0.054	0.047	0.045	0.024
$\hat{\lambda}^{(2)}$	OLS	mean	0.113	0.200	0.292	0.391	0.480	0.590	0.691	0.784	0.889
		$\sqrt{\text{MSE}}$	0.064	0.076	0.076	0.073	0.075	0.062	0.054	0.043	0.027

Table D.3: **Regression results for coefficient of tail dependence estimation of Clayton copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 100 simulations of 1000 pairs of random variables drawn from a Clayton copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.182	0.251	0.332	0.411	0.493	0.594	0.692	0.794	0.895
		$\sqrt{\text{MSE}}$	0.093	0.067	0.062	0.048	0.048	0.045	0.041	0.032	0.018
$\hat{\lambda}^{(1)}$	OLS	mean	0.166	0.234	0.321	0.404	0.494	0.588	0.691	0.787	0.892
		$\sqrt{\text{MSE}}$	0.085	0.067	0.064	0.049	0.056	0.047	0.040	0.033	0.018
$\hat{\lambda}^{(2)}$	WLS	mean	0.188	0.255	0.336	0.414	0.496	0.596	0.695	0.797	0.896
		$\sqrt{\text{MSE}}$	0.098	0.071	0.063	0.048	0.048	0.045	0.041	0.033	0.019
$\hat{\lambda}^{(2)}$	OLS	mean	0.170	0.237	0.323	0.406	0.495	0.590	0.692	0.788	0.893
		$\sqrt{\text{MSE}}$	0.087	0.068	0.065	0.049	0.056	0.047	0.040	0.033	0.018

Table D.4: **Regression results for coeffient of tail dependence estimation of Gumbel copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 100 simulations of 1000 pairs of random variables drawn from a Gumbel copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.102	0.191	0.303	0.395	0.499	0.593	0.695	0.797	0.898
		$\sqrt{\text{MSE}}$	0.038	0.054	0.053	0.052	0.044	0.046	0.038	0.022	0.016
$\hat{\lambda}^{(1)}$	OLS	mean	0.101	0.192	0.302	0.389	0.498	0.589	0.692	0.792	0.896
		$\sqrt{\text{MSE}}$	0.043	0.057	0.060	0.058	0.050	0.049	0.039	0.025	0.016
$\hat{\lambda}^{(2)}$	WLS	mean	0.110	0.197	0.307	0.398	0.501	0.594	0.695	0.797	0.899
		$\sqrt{\text{MSE}}$	0.038	0.052	0.052	0.052	0.044	0.046	0.039	0.023	0.017
$\hat{\lambda}^{(2)}$	OLS	mean	0.106	0.196	0.305	0.390	0.499	0.590	0.692	0.793	0.896
		$\sqrt{\text{MSE}}$	0.043	0.056	0.059	0.058	0.050	0.049	0.040	0.025	0.016

Table D.5: **Regression results for coeffient of tail dependence estimation of Clayton copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 100 simulations of 10000 pairs of random variables drawn from a Clayton copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.097	0.200	0.291	0.401	0.496	0.595	0.702	0.799	0.899
		$\sqrt{\text{MSE}}$	0.019	0.025	0.030	0.027	0.033	0.025	0.018	0.016	0.008
$\hat{\lambda}^{(1)}$	OLS	mean	0.094	0.201	0.292	0.401	0.497	0.591	0.699	0.797	0.898
		$\sqrt{\text{MSE}}$	0.022	0.030	0.039	0.033	0.042	0.030	0.022	0.016	0.008
$\hat{\lambda}^{(2)}$	WLS	mean	0.098	0.201	0.291	0.401	0.496	0.595	0.702	0.799	0.900
		$\sqrt{\text{MSE}}$	0.019	0.025	0.030	0.027	0.033	0.025	0.018	0.016	0.008
$\hat{\lambda}^{(2)}$	OLS	mean	0.095	0.202	0.293	0.401	0.497	0.591	0.699	0.797	0.898
		$\sqrt{\text{MSE}}$	0.022	0.030	0.039	0.033	0.042	0.030	0.022	0.016	0.008

Table D.6: **Regression results for coeffient of tail dependence estimation of Gumbel copula.**

This table shows the means and the mean squared errors (MSE) for the WLS and OLS regression of the two tail dependence estimators. This is based on 100 simulations of 10000 pairs of random variables drawn from a Gumbel copula with tail dependence coefficient $\lambda = 0.1, 0.2, \cdots, 0.9$.

Estimator	Regression type	λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\hat{\lambda}^{(1)}$	WLS	mean	0.158	0.230	0.317	0.406	0.496	0.599	0.705	0.798	0.903
		$\sqrt{\text{MSE}}$	0.065	0.039	0.033	0.033	0.027	0.028	0.022	0.014	0.010
$\hat{\lambda}^{(1)}$	OLS	mean	0.150	0.224	0.316	0.404	0.496	0.597	0.701	0.798	0.900
		$\sqrt{\text{MSE}}$	0.060	0.038	0.035	0.035	0.030	0.027	0.022	0.013	0.008
$\hat{\lambda}^{(2)}$	WLS	mean	0.159	0.231	0.318	0.407	0.497	0.599	0.705	0.798	0.903
		$\sqrt{\text{MSE}}$	0.066	0.040	0.033	0.033	0.027	0.028	0.022	0.014	0.010
$\hat{\lambda}^{(2)}$	OLS	mean	0.151	0.224	0.316	0.405	0.496	0.597	0.701	0.798	0.901
		$\sqrt{\text{MSE}}$	0.060	0.038	0.035	0.035	0.030	0.027	0.022	0.013	0.008

For both estimators we see that both regressions works better for large tail indices. This effect is stronger for a sample from a Clayton copula.

D.2. Copula estimation

In section 4.11.4 we showed the performance of the copula estimation introduced in section 4.9.3 on samples from a Gumbel copula. Here, we do the same for a Clayton copula. In Table D.7 we compare the dependence measures of the sample from the Clayton copula and the sample from our copula estimation procedure, based on 50 samples for each TDC.

Table D.7: **Comparison dependence measures of scenarios and historical data.**

This table shows important dependence measures of our 100 generated scenarios of 1000 observations and compares those to the dependence measures of the underlying sample, which are four samples of 1000 observations generated from a Clayton copula for $\lambda = 0.2, 0.4, 0.6, 0.8$.

TDC	Theoretical	0.2	0.4	0.6	0.8
	Sample	0.250	0.395	0.509	0.770
	Scenarios	0.233	0.382	0.501	0.731
	Rejection percentage of scenario	14%	32%	38%	30%
Tau	Theoretical	0.177	0.274	0.404	0.608
	Sample	0.161	0.258	0.395	0.617
	Scenarios	0.162	0.258	0.395	0.616
	MSE	0.021	0.021	0.019	0.014

The differences in the dependence measures in Table D.1 are small, but we underestimate the TDC slightly. This underestimation depends on the grid size and we therefore compute the MSEs in the Kendall's tau and TDC for different grid sizes and show these in Table D.8.

Table D.8: **Comparison MSEs of scenarios for different grid sizes.**

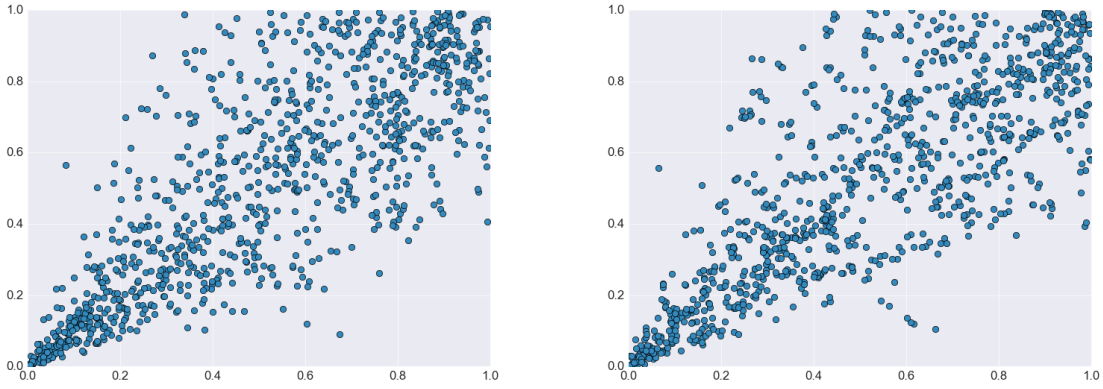
This table shows the average MSE of the scenarios $\hat{\tau}$ and λ_U for a Clayton copula as a function of the grid size. The size of the historical and scenario data is 1000 and the grid sizes are 10, 20 and 40.

Amount of grid points	10	20	40
TDC	0.156	0.039	0.019
tau	0.026	0.018	0.018

Table D.8 shows that the MSEs of the dependence measures are smaller for finer grids. An example of a scenario generated from the copula is given in Figure D.1 and compared to the original sample generated from the Clayton copula. We have a grid of 40 equally spaced points.

Figure D.1: **Generated historical sample and scenario sample.**

This figure gives the historical sample on which we base our copula on the left and a scenario generated from this historical sample on the right. The historical sample is generated from a Clayton copula with $\lambda_L = 0.8$



(a) Sample from Clayton copula

(b) Sample from copula estimated on left sample

In Figure fi:copula historical and scenario clayton the estimated Kendall's taus are simular: $\tau = 0.613$ vs $\tau = 0.611$. The TDC estimates are given by $\lambda_L = 0.802$ and $\lambda_L = 0.784$.

E

Methods

E.1. Ranking method

In section 4.3.2 we introduce the Kendall's tau and Spearman's rho, that both depend on the ranking of the data. Due to truncation in the data we can have degenerate pairs and we need to find a method to rank these. Ordinal and average ranking seem to be accurate options to deal with degeneracy and are explained below. We come up with a third way of ranking, namely reverse ordinal ranking, which might be counter-intuitive but may be useful in some cases. This method ranks completely opposite to the ordinal ranking method and is also explained below.
(1) Ordinal ranking: Give the smallest rank to the first element, the second smallest rank to the second element and so forth.
(2) Average ranking: Give all these observations the average rank of the above method.
(3) Reverse ordinal ranking: Give the largest rank to the first element, the largest but one rank to the second element and so forth.

When ranking 2-dimensional data we can use combinations of these methods on the two dimensions. We illustrate three possible ranking methods in Example E.1.

Example E.1. *Suppose we have the following observations:* $(10,8)$, $(2,4)$, $(2,4)$, *and* $(1,1)$. *Then we obtain the ranks in Table E.1 by using the ordinal, average and reverse ordinal ranking methods.*

Table E.1: **Different ranking methods.**

This figure shows the results of three ranking methods applied to the observations in Example E.1. The three methods are: ordinal ranking on both dimensions (1), average ranking on both dimensions (2), and ordinal ranking on the first dimension and reverse ordinal ranking on the second dimension (3).

Observations	$R^{(1)}$	$R^{(2)}$	$R^{(3)}$
$(10,8)$	$(4,4)$	$(4,4)$	$(4,4)$
$(2,4)$	$(2,2)$	$(2.5,2.5)$	$(2,3)$
$(2,4)$	$(3,3)$	$(2.5,2.5)$	$(3,2)$
$(1,1)$	$(1,1)$	$(1,1)$	$(1,1)$

Suppose we have fully correlated variables, with one equal observation $X^i = X^j$. Then average ranking gives us that $R_1^i = R_1^j = R_2^i = R_2^j$, such that ρ_S still gives 1. The pair is however neither concordant nor discordant, so we will have $\tau < 1$. Using ordinal ranking, we find concordance for these pairs and thus $\tau = 1$, also $\rho_S = 1$ still holds. The reverse ordinal ranking gives an even smaller τ than the average ranking and also gives $\rho_S < 1$ and should therefore not be used for fully correlated variables. Hence, ordinal ranking is the best for the fully correlated case. Using other ranking methods namely underestimates the dependence when we have degeneracy. In case of anticorrelation it is the other way around and in this case we should use reverse ordinal ranking. The best estimator, thus, depends on the data.

To decide which ranking method to use, we can simply make an initial guess of the dependence using the average ranking and then use the right ranking method. When the initial guess gives $\rho_S > 0$ we will use

ordinal ranking and when $\rho_S < 0$ we will use reverse ordinal ranking (in case of $\rho_S = 0$ we will stay with the initial guess).

A similar problem with ranking arises when from a pair of observations one of the two vector elements is equal and the other elements differ (e.g. Example E.1 with $(2,4)$ and $(2,4.1)$ instead of $(2,4)$ and $(2,4)$). In this case we propose a similar ranking method as the one before. We would first rank using average ranking, which would give $(2.5,2.5)$ for both of these observations. When this ranking gives $\rho_S > 0$, which is the case here, we will give the one with the larger rank for the second element the larger rank for the first element. The ranks for (2.4) and $(2,4.1)$ will then become $(2,2)$ and $(3,3)$ respectively. When $\rho_S < 0$ we would rank the other way around and obtain ranks $(3,2)$ and $(2,3)$.

We illustrate the method in the following example for further clarification.

Example E.2. *Suppose we have the following observations:* $(10,8)$, $(2,4.1)$, $(2,4)$, $(2,4)$, *and* $(1,1)$. *Then we would get initial guess* $R_{(0)}$ *using the average ranking. Next, based on the* ρ_S *calculation, we will choose the final ranking method and get* $R_{(1)}$.

Table E.2: **Ranks of our observations.**

This figure show the initial ranks using the average ranking method and the final ranks using the ordinal ranking method, which is used since the initial ranking gives $\rho_S > 0$.

Observations	$R_{(0)}$	$R_{(1)}$
$(10,8)$	$(5,5)$	$(5,5)$
$(2,4.1)$	$(3,4)$	$(4,4)$
$(2,4)$	$(3,2.5)$	$(2,2)$
$(2,4)$	$(3,2.5)$	$(3,3)$
$(1,1)$	$(1,1)$	$(1,1)$

We can immediately see from the ranks that we obtain $\tau = \rho_S = 1$ by using $R^{(1)}$ in this example, which is what we want, since we do not want to exclude full dependence.

E.2. Estimation of parameters for parametric copulas

Suppose we have chosen a parametric model for the copula, which describes the multivariate behaviour of X and Y. Before we can proceed with the maximum likelihood estimation (MLE) for the model parameter θ, we need the expression for the multivariate density function of X and Y. The likelihood for a single observation when the samples are independent is given in equation E.1, see Jaworski et al. (2010).

$$L(x_1,\cdots,x_n) = f(x_1,\cdots,x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} C(F_1(x_1),\cdots,F_n(x_n);\theta) = c(F_1(x_1),\cdots,F_n(x_n);\theta) \prod_{i=1}^{n} f_i(x_i), \tag{E.1}$$

where c is the copula density, recall equation (4.2). The corresponding log-likelihood is then given by:

$$l(x_1,\cdots,x_n) = \log c(F_1(x_1),\cdots,F_n(x_n);\theta) + \sum_{i=1}^{n} \log f_i(x_i),$$

where c is the copula density given by equation (4.2). For independent observations $\{X^1,\cdots,X^m\} \in \mathbb{R}^n$ we get the following likelihood:

$$l(X^1,\cdots,X^m) = \sum_{j=1}^{m} \left\{ \log c\left(F_1\left(X_1^j\right),\cdots,F_n\left(X_n^j\right);\theta\right) + \sum_{i=1}^{n} \log f_i\left(X_i^j\right) \right\}. \tag{E.2}$$

Minimizing this log-likelihood gives us the parameters of the model. We can therefore obtain the MLE estimate by solving equation (E.3), see Jaworski et al. (2010).

$$\frac{\partial}{\partial\theta} l(X^1,\cdots,X^m) = \frac{\partial}{\partial\theta} \sum_{j=1}^{m} \left\{ \log c\left(F_1\left(X_1^j\right),\cdots,F_n\left(X_n^j\right);\theta\right) + \sum_{i=1}^{n} \log f_i\left(X_i^j\right) \right\} = 0. \tag{E.3}$$

Bibliography

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and economics*, 44, no. 2, pp. 182–198, 2009.

H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, 19, no. 6, pp. 716–723, 1974.

C.D. Aliprantis and O. Burkinshaw. *Principles of Real Analysis*. Academic Press, 1998.

R. Aloui, M.S.B. Aïssa, and D.K. Nguyen. Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking and Finance*, 35, no. 1, pp. 130–141, 2011.

D.G. Baur and T.K. McDermott. Is gold a safe haven? International evidence. *Journal of Banking and Finance*, 35, no. 8, pp. 1886–1898, 2010.

J. Beirlant, G. Dierckx, Y. Goegebeur, and G. Matthys. Tail index estimation and an exponential regression model. *Extremes*, 2, no. 2, pp. 177–200, 1999.

F. Caeiro and M.I. Gomes. Threshold selection in extreme value analysis. *Extreme Value Modeling and Risk Analysis Methods and Applications*, pages 69–86, 2016.

S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2, pp. 339–365, 1999.

F. Durante, J. Fernández-Sánchez, J.J. Quesada-Molina, and M. Úbeda-Flores. Copulas with given values on the tails. *International Journal of Approximate Reasoning*, 85, pp. 59–67, 2017.

P. Embrechts, F. Lindskog, and A. Mcneil. Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384, 2003.

L. Fahrmeir, T. Kneib, S. Lang, and B. Marx. *Regression*. Springer, 2007.

M. Falk and R. Michel. Testing for tail independence in extreme value models. *Annals of the Institute of Statistical Mathematics*, 58, pp. 261–290, 2006.

G. Frahm, M. Junker, and R. Schmidt. Estimating the tail-dependence coefficient: properties and pitfalls. *Insurance: Mathematics and Economics*, 37, pp. 80–100, 2005.

C. Genest and J. Neslehova. A primer on copulas for count data. *ASTIN Bulletin: The Journal of the IAA*, 37, no. 2, pp. 475–515, 2007.

M.I. Gomes and M.J. Martins. "asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. *Extremes*, 5, no. 1, pp. 5–31, 2002.

L. de Haan and A. Ferreira. *Extreme value theory*. Springer, 2006.

L. de Haan and S. Resnick. Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of Probability*, pages 97–124, 1996.

P. Hall and A. Guillou. A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, no. 2, pp. 293–305, 2001.

P. Hall and A.H. Welsh. Adaptive estimates of parameters of regular variation. *The Annals of Statistics*, 13, no. 1, pp. 331–341, 1985.

N. Hyung and C.G. de Vries. Portfolio selection with heavy tails. *Journal of Empirical Finance*, 14, no. 3, pp. 383–400, 2007.

P. Jaworski, F. Durante, W.K. Hardle, and T. Rychlik. Copula theory and its applications. 2010.

H. Joe, H. Li, and A.K. Nikoloulopoulos. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101, no. 1, pp. 252–270, 2010.

I.T. Jolliffe. Principal component analysis and factor analysis. In *Principal component analysis*, pages 115–128. Springer, 1986.

D.X. Li. On default correlation: A copula function approach. *The Journal of Fixed Income*, 9, no. 4, pp. 43–54, 2000.

Y. Malevergne and D. Sornette. Testing the Gaussian copula hypothesis for financial assets dependences. *Quantatative Finance*, 3, no. 4, pp. 231–250, 2003.

Y. de Malevergne and D. Sornette. *Extreme Financial Risks*. Springer, 2006.

R.B. Nelsen. *An Introduction to Copulas*. Springer, 1999.

R. Reiss and M. Frick. Expansions of multivariate Pickands densities and testing the tail dependence. *Journal of Multivariate Analysis*, 100, pp. 1168–1181, 2009.

F. Salmon. The formula that killed wall street. *Significance*, 9, no. 1, pp. 16–20, 2012.

R. Schmidt. Tail dependence. *Statistical tools in finance and insurance. Springer, New York*, 2005.

V. Schmitz. *Copulas and stochastic processes*. PhD thesis, Bibliothek der RWTH Aachen, 2003.

B. Schweizer and A. Sklar. Operations on distribution functions not derivables from operations on random variables. *Studia Mathematica*, 52, pp. 43–52, 1974.

B. Schweizer and E.F. Wolff. On nonparametric measures of dependence for random variables. *The Annals of Statistics*, 9, no. 4, pp. 879–885, 1981.

R.H. Shumway and D.S. Stoffer. *Time series analysis and its applications: with R examples*. Springer Science & Business Media, 2010.

A. Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9, no. 6, pp. 449–460, 1973.

C. Starica. Multivariate extremes for models with constant conditional correlations. *Journal of Empirical Finance*, 6, pp. 515–553, 1999.

H. Steehouwer. A frequency domain methodology for time series modelling. In *Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds*, pages 280–324. Springer, 2010.

J. Stewart. *Calculus*. Cengage Learning, 2011.

P.R. Tadikamalla. A look at the burr and related distributions. *International Statistical Review/Revue Internationale de Statistique*, pages 337–344, 1980.

B. Vandewalle, J. Beirlant, and M. Hubert. A robust estimator of the tail index based on an exponential regression model. *Theory and applications of recent robust methods*, pages 367–376, 2004.

G.A. Young and R.L. Smith. *Essentials of statistical inference*, volume 16. Cambridge University Press, 2005.