

# Portfolio Risk Management using Extreme Value Theory and Vine Copulas

## MASTER THESIS

**Written by:** Jaron Wit - 400854 <sup>1</sup>

Supervisors:

X. Leng - Erasmus University Rotterdam  
P. Bosschaart - KPMG

Erasmus School of Economics, Erasmus University Rotterdam

KPMG Financial Risk Management, Amstelveen

May 21, 2018

### Abstract

We estimate the one- and ten-day-ahead Value-at-Risk and the Expected Shortfall of a portfolio by using marginal modeling and two different dependence models. The marginal modeling consists of two steps. Firstly the returns are filtered by an asymmetric Markov-Switching ARMA-GARCH model (ARMA-MS-GARCH-GJR) and secondly the tails of the filtered returns are estimated by the Generalized Pareto Distribution. We fit two different joint distributions on these transformed returns: Vine Copulas and a distribution based on Multivariate Extreme Value Theory (MEVT). These two models will be compared with the t copula by backtesting the Value-at-Risk (VaR) and Expected Shortfall (ES) in two periods: during the Financial Crisis in 2008 and during a stable period. We show from the backtesting that all joint distributions are capable of providing accurate one-day-ahead VaR and ES forecasts, while the ten-day-ahead forecasts during the Financial Crisis are not accurate. It follows that during periods of financial distress the models with fatter tails are in favor, such as the t copula and the Vine Copula, while in stable periods these models tend to overestimate the Value-at-Risk and the Expected Shortfall slightly. In stable periods the model based on MEVT is preferred.

**Keywords:** *Portfolio VaR and ES, Vine Copula, Multivariate Extreme Value Theory, Markov-Switching ARMA-GARCH-GJR, Generalized Pareto Distribution, Backtesting*

---

<sup>1</sup>Master Student Quantitative Finance at the Erasmus School of Economics, Erasmus University Rotterdam.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Marginal Modeling . . . . .	1
1.2	Multivariate Modeling . . . . .	2
1.2.1	Vine Copulas . . . . .	3
1.2.2	Multivariate Extreme Value Theory . . . . .	3
<b>2</b>	<b>Literature</b>	<b>4</b>
<b>3</b>	<b>Data</b>	<b>5</b>
3.1	Data Transformation . . . . .	6
3.2	Indices Dataset . . . . .	6
<b>4</b>	<b>Methodology</b>	<b>8</b>
4.1	Risk Measures . . . . .	8
4.1.1	Value-at-Risk (VaR) . . . . .	9
4.1.2	Expected Shortfall (ES) . . . . .	9
4.2	Dependence Measures . . . . .	9
4.2.1	Kendall's $\tau$ ( $\tau$ ) . . . . .	10
4.2.2	Tail Dependence . . . . .	10
4.3	GARCH modeling . . . . .	10
4.3.1	ARMA-GARCH-GJR . . . . .	11
4.3.2	ARMA-MS-GARCH-GJR . . . . .	12
4.3.3	Ljung Box test . . . . .	13
4.3.4	Augmented Dickey-Fuller test . . . . .	13
4.4	Semi-Parametric modeling and Extreme Value Theory (EVT) . . . . .	13
4.5	Copulas . . . . .	15
4.5.1	The Concept of Copulas . . . . .	15
4.5.2	Copula Families . . . . .	16
4.5.3	Copula Estimation . . . . .	17
4.5.4	Copula Selection . . . . .	17
4.5.5	Independence Test . . . . .	18
4.6	Pair-Copula Constructions . . . . .	18
4.6.1	Pair-Copula Constructions (PCC) . . . . .	18
4.6.2	Regular Vines (R-Vines) . . . . .	20
4.6.3	Estimating Copulas . . . . .	22
4.6.4	Sampling from Vine Copulas . . . . .	23
4.6.5	Sequential Method . . . . .	23
4.6.6	Model Comparison of R-Vines . . . . .	24
4.7	Multivariate Extreme Value Theory . . . . .	25
4.7.1	MEVT Model . . . . .	25
4.7.2	MEVT estimation . . . . .	26
4.7.3	MEVT Sampling . . . . .	26
4.8	Backtesting . . . . .	28
4.8.1	POF-test . . . . .	29
4.8.2	ES Zero Mean Test . . . . .	29

<b>5 Results</b>	<b>29</b>
5.1 Marginal Results . . . . .	30
5.1.1 ARMA-MS-GARCH-GJR . . . . .	30
5.1.2 Marginal Extreme Value Theory . . . . .	33
5.2 Multivariate Results . . . . .	33
5.3 Vine Copula results . . . . .	33
5.4 MEVT results . . . . .	33
5.5 Backtesting Results . . . . .	34
<b>6 Conclusion</b>	<b>43</b>
<b>A Derivations</b>	<b>47</b>
A.1 Hamilton Filter . . . . .	47
A.1.1 Prediction Step . . . . .	47
A.1.2 Updating Step . . . . .	47
A.1.3 Estimation of Hamilton Filter . . . . .	48
A.2 Marginal GPD . . . . .	48
A.3 Five dimensional C-Vine and D-Vine . . . . .	48
<b>B Tables</b>	<b>49</b>
<b>C Figures</b>	<b>52</b>

# 1 Introduction

Financial risk management is the process of quantifying and dealing with the uncertainty of bad outcomes in the financial world. There are different types of financial risks, such as market risk, credit risk and operational risk. In this research the risk of losses when taking certain positions in the financial market, or the so-called market risk of a portfolio, is considered. Risk measures are used to determine the riskiness of financial instruments by quantifying the risk. Risk measures therefore play a central role in the financial risk management. The behavior of the profit or loss distribution <sup>2</sup> is often used in order to estimate risk measures. There are two risk measures which are in practice extensively used: the Value-at-Risk (VaR) and the Expected Shortfall (ES) (McNeil et al. 2005). If we consider losses, then the  $\text{VaR}_\alpha$  corresponds to the loss which has a probability  $1 - \alpha$  of being exceeded by a larger loss. The  $\text{VaR}_\alpha$  can be seen as the  $\alpha$  quantile in a loss distribution. VaR reporting is part of the capital-adequacy framework of Basel II. It is required that the ten-day ahead  $\text{VaR}_{99\%}$  is reported. This ten-day  $\text{VaR}_{99\%}$  corresponds to the value that a cumulative ten-day loss will only exceed one in the hundred times (1%). However, this value tells nothing about the size of the potential loss when it is exceeded. The ES does. The ES is more sensitive to the shape of the tail. The  $\text{ES}_{99\%}$  is the average of all the losses greater than the  $\text{VaR}_{99\%}$ . In this research the target is to estimate the VaR and the ES not only for a single financial instrument, but for a certain position in multiple financial instruments or a portfolio. This adds another dimension to the problem. It is well-known that financial instruments often show strong dependencies which need to be accounted for when constructing a portfolio VaR or ES. As our focus mainly lies on the extreme quantiles of our returns we need to incorporate the joint tail behavior of the marginals. To model the dependencies we will use two multivariate models. The first model is based on the Vine Copula theory from Bedford & Cooke (2002) and Dißmann (2010). The second model is based on the Multivariate Extreme Value Theory (MEVT) from Hilal et al. (2014). Both models require data that is independent and identically distributed as marginal input. This requires filtering in the first place. In the end we will compare the performance of both different multivariate approaches by comparing the backtest results of the VaR and the ES.

## 1.1 Marginal Modeling

Before estimating the joint distribution we filter the returns. With our filter we aim to transform the data into i.i.d. observations, which is required when modeling the joint distributions. It is well-known that financial time series exhibit several stylized facts. McNeil et al. (2005) summarizes these stylized facts into six bullet points <sup>3</sup>. Given the fact that many time series we will be using may exhibit several of these characteristics, we choose our marginal models in such a way that these characteristics are captured. There are several ways to model the marginals. Firstly there are non-parametric methods. These non-parametric methods do not take any financial return characteristics into account. Historical simulation is a common used non-parametric methods to model marginal distributions. Secondly there are parametric methods. Examples of parametric methods include GARCH, extreme value theory and kernel fitting, but there are many more. The benefits of using non-parametric methods are mainly their simplicity and the fact that no distributional assumption is required. However there are some drawbacks of non-parametric marginal modeling. Firstly there is no extrapolation and interpolation possible when using a non-parametric method like historical simulation. Furthermore empirical quantile estimations tend to be very volatile when large observations enter the sample (McNeil & Frey 2000). It is found that the parametric methods outperform the non-parametric methods. In order words, incorporating stylized facts in our models is beneficial. We can use GARCH for example to account for volatility clustering and extreme value theory to model the tail behavior.

---

<sup>2</sup>When returns are multiplied by minus one and these values are used to construct a distribution, then the distribution becomes a loss distribution instead of a profit distribution.

<sup>3</sup>The following facts are often considered to be stylized in financial time series: (1) Return series are not i.i.d. although they show little serial correlation; (2) Series of absolute or squared returns show profound serial correlation; (3) Conditional expected returns are close to zero; (4) Volatility appears to vary over time; (5) Return series are leptokurtic or heavy-tailed; (6) Extreme returns appear in clusters (McNeil et al. 2005, page. 117).

The two step method from McNeil & Frey (2000), the method that we use, has become a widely-used basis for VaR estimation. Combining GARCH with Generalized Pareto Distribution (GPD) for the tails has proven to be powerful when estimating the VaR and ES (McNeil & Frey 2000). The GARCH model is used to filter out possible volatility clustering, while extreme value theory is used to capture the tail behavior by modeling the upper- and the lower-tails separately using the GPD. This allows us to model heavy tails and asymmetric tail behavior. We not only choose to use this two step method, but we also extent it. To adsorb any autocorrelation in the returns we choose to extent the GARCH model with an ARMA model. Besides volatility clustering we want to incorporate the leverage effect. This effect arises from the observation that in many financial time series negative shocks in the very recent past tend to have a stronger effect on the variance than positive shocks. We therefore choose to add an asymmetric component to the GARCH model. We therefore choose to use the GARCH-GJR model from Glosten et al. (1993). Since Tang & Gau (2004) show that the volatility process may differ across different regimes, we choose make the ARMA-GARCH-GJR Markov-Switching over time to govern the return by a non-observed Markov chain. Tang & Gau (2004) show that this Markov-Switching GARCH variant is beneficial when it comes to VaR estimation. The model that we will use will be called the ARMA-MS-GARCH-GJR model. We use this model to obtain a filtered time series that is very close to i.i.d. which is required prior to the multivariate modeling. The filtered series will be tested on i.i.d. by using a Ljung-Box test. Furthermore the Augmented Dickey-Fuller test will be conducted in order to check for stationarity.

Once the returns are filtered, we will model these filtered returns in a semi-parametric way. This corresponds to the next step (McNeil & Frey 2000). With semi-parametric modeling we model both the lower and upper tails by using different extreme value distribution functions, while we model the middle part of the distribution by the empirical distribution. Modeling the upper and lower tails using extreme value theory (EVT) allows for interpolation and extrapolation in the regions where the amount of data are scarce. Due to the fact that the tails of the standardized data may behave in a different manner, both tails will be fitted into separate models. To use data more efficiently, we choose to base the extreme value density function on exceedances over thresholds instead of on Block Maxima, see McNeil et al. (2005). The distribution that corresponds to exceedances over thresholds is the GPD. Choosing the right threshold is a balance between bias and variance. Allowing a huge number of observations in the tail results to smaller variance and introduces a larger bias. On the other hand, allowing fewer observations will result into larger variance and smaller bias. The thresholds are chosen using Hill plots and an automated eye-ball test from Danielsson et al. (2016).

## 1.2 Multivariate Modeling

Once all the returns are filtered in the first step, the filtered returns will be transformed to the standard uniform scale by the corresponding fitted semi-parametric marginal distribution. We choose to do so, since both dependence models require (close to) i.i.d. data in the standard uniform scale. The joint distributions are used to model the dependence between the components of our portfolio.

In order to show the importance of multivariate modeling in the area of risk management we use an example from during the 2008 Financial Crisis. In October 2007 the tranches of Collateralized Debt Obligations (CDOs) that are considered to be the risk safest (AAA-tranche) started to fall and turned out to be strongly correlated<sup>4</sup>. The domino-like collapse of the network of promises that made up the CDO market could be seen as one of the causes of the 2008 financial crisis. The Financial Times published an article "Recipe for Disaster: The Formula That Killed Wall Street", which shed light on the dependence structures that were used prior to the crisis for Collateralized Debt Obligations. It points out that the formulas that were used (Gaussian Copulas) did not account for the strong dependence of CDOs during financial distress. This example underlines the importance of proper dependence modeling, mainly during times of financial distress. In this research we use two methods to model the joint behavior of a set of financial return series.

---

<sup>4</sup>The Financial Crisis Inquiry Report, 2011, p.129

### 1.2.1 Vine Copulas

The first method to model the joint distribution is based on copula theory. Copulas allow us to combine complex marginal models with different dependence structures. In other words the marginal distributions are individually differently specified, completely apart from the dependence function. This property make copulas desired dependence models. The dependence structure between financial data is characterized by a non-linear relationship (Embrechts, McNeil & Straumann 1999). Copulas allow a non-linear relationship between variables. The construction of higher-dimensional copulas (often when there are more than two dimensions) is recognized as a difficult problem (Embrechts et al. (2003) and Savu & Trede (2006)). The Gaussian and the t copulas, which are both elliptical, can easily be extended to higher dimensions, while many other copulas cannot. Estimation of many multivariate copulas are characterized by long computation time, see e.g. Embrechts et al. (2003), Savu & Trede (2006) and Czado et al. (2010)(page. 93-109). Throughout the history many bivariate copulas have been developed, but the number of multivariate ones is rather limited (Gruber & Czado 2015). Besides the difficulties that arise when constructing multivariate copulas, Hilal et al. (2011) elaborate on the fact that multivariate copulas often have hard and inflexible dependence conditions. Regular Vine copulas provide a solution to this problem by offering a method that can construct a multivariate density function as a product of bivariate conditional and unconditional copula densities (Joe (1996) and Bedford & Cooke (2002)). A multivariate density is split up into a cascade of different conditional and unconditional bivariate copulas (pair-copulas). In other words these pair-copulas will be used as simple building blocks to construct higher dimensional dependence structures. There are however a lot of different ways to split up a multivariate system into pair-copulas. A remaining question is how to choose this structure. Arbitrarily, one often use C or D-vine structures, because of the graphical simplicity of these structures. Following Dißmann (2010) Graph theory and maximum spanning trees can be used to adsorb most of the dependence in the first pair-copulas. Dißmann (2010) provide us several reasons why this is considered to be beneficial. Firstly, it is important to hold in mind that most copulas can in general model independence and close to independence pretty well. Secondly, the first copula pairs have the most influence on the rest of the consecutive pairs. Thirdly, it is natural to assume that selecting the variables with the strongest dependence in the first trees will adsorb the dependence in later trees. In other words, choosing a certain R-Vine using strongest dependencies in the first tree, will result to many weak dependence in later trees. So, a small number of bivariate copulas model stronger dependencies, while a large number is used to model pairs that are close to independence, therefore there is less room for estimation error. Dißmann (2010) shows that the sample fit of these Vine copulas is generally better in terms of likelihood than Gaussian and t copulas (Aas et al. (2009) and Dißmann (2010)).

### 1.2.2 Multivariate Extreme Value Theory

The second method that is used to model the dependence structure is based on the multivariate extreme value theory from Hilal et al. (2014). Again this model requires i.i.d. marginal data as input, which is obtained by the filtering in the marginal modeling step. This multivariate extreme value distribution has a non-linear dependence structure. According to Hilal et al. (2011) there are mainly two advantages of using this approach. Firstly, there is besides copulas virtually no model that is capable of handling extreme value distributions in more than two dimensions from a practical point of view. Secondly many extreme value distributions assume tail dependence once the assumption of independent variables is rejected. If variables are assumed to be asymptotic dependent while in reality they are not, then this will lead to an overestimation of the probability of joint extreme events (Ledford & Tawn 1996). The model of Hilal et al. (2014) overcomes the issues stressed out above. The model captures the asymmetric behavior of extreme negative and positive returns. Once the marginals are modeled, the data on standard uniform scale are transformed using the Laplace transformation. The dependence modeling is based on the distribution of a certain asset of our portfolio, given the fact that another asset is considered to be extreme. All the possible different possible "pairs" are estimated using Maximum Likelihood estimation. In the case that both observations of assets are considered to be non-extreme, the distribution is modeled empirically.

The simulated data from both multivariate distributions are scaled from zero to one. This data will be transformed by the inverse of the corresponding estimated marginal semi-parametric models to obtain filtered residuals. These filtered residuals are used in the ARMA-MS-GARCH-GJR model to obtain simulated returns.

We show in this research that an unobserved Markov-chain with two different states clearly distinguishes between turbulent and stable financial periods. The turbulent financial periods are characterized by a lower persistence of historical variance. On the other hand during turbulent periods the ARMA-GARCH-GJR models are more strongly exposed to the leverage effect. We show from the backtesting that all joint distribution are capable of providing accurate one-day-ahead VaR and ES forecasts, while the ten-day-ahead forecasts during the Financial Crisis are not accurate. It follows that during periods of financial distress the models with fatter tails are in favor, such as the t copula and the Vine Copula, while in stable periods these models tend to overestimate the Value-at-Risk and the Expected Shortfall slightly. In stable periods the model based on MEVT is preferred. The remainder of this paper will be as follows. Section 2 describes the findings in literature for both the marginal and the joint modeling. We describe our data transformations in Section 3 along with some descriptive statistics. In Section 4 we provide the methodology for both the marginal and the joint models. We will discuss the results in section 5 and finally draw conclusions in Section 6.

## 2 Literature

In this section we give an overview of relevant literature. This overview is twofold: in the first part we provide relevant literature on the marginal modeling and in the second part literature on the joint modeling, which contains the copula modeling and the modeling based on MEVT.

The marginal modeling is based on the two step approach from McNeil & Frey (2000). McNeil & Frey (2000) proposed to combine two powerful models that account for several different characteristics that financial time series exhibit. Firstly the data is filtered by using GARCH models. These type of models are commonly used for estimating the VaR or the ES. Secondly the filtered data is modeled semi-parametrically. The tails of the filtered data are modeled by the GPD apart from the main body, which is modeled non-parametrically by using the empirical density function. Instead of using an ordinary GARCH model as in McNeil & Frey (2000), we choose to extent it. Mabrouk & Chaker (2011) give an overview of different GARCH models and show that inclusion of asymmetry in GARCH improves the performance. Furthermore Mabrouk & Chaker (2011) show that a GARCH model with an ARMA extension performs well when forecasting the VaR. Therefore we use a GARCH model with both an ARMA and a GJR extension, resulting to ARMA-GARCH-GJR model. From this point we deviate slightly from the closely related work of Dikmann (2010) and Hilal et al. (2014), who use a GARCH model with only a ARMA extension. Since Tang & Gau (2004) show that the volatility process may differ across different regimes, we choose make the ARMA-GARCH-GJR Markov-Switching over time by governing the return series by a non-observed Markov chain. Tang & Gau (2004) show that this Markov-Switching GARCH variant performs well in terms of VaR estimation. We refer to this final model as the ARMA-MS-GARCH-GJR model.

The tail modeling in the second step of McNeil & Frey (2000) approach has been extensively studied. Roughly speaking there are two different approaches for putting extreme value theory into practice. The first approach is the most traditional one. It is based on block maxima theory. It has extensively studied by Embrechts et al. (1997), Longin (1997) and McNeil (1998). In this approach the largest observations from large samples of i.i.d distributed observations are collected in blocks. Longin (1997) and Embrechts et al. (1997) constructed models based on block maxima disregarding the fact that volatility tends to cluster over time <sup>5</sup>. A more modern group of models is based on the exceedances over thresholds. This type of model is proposed by Embrechts, Resnick & Samorodnitsky (1999). It has extensively been used in modeling losses in various ways, such as: financial losses (Embrechts, McNeil & Straumann (1999) and McNeil & Frey (2000)), windstorm losses (Rootzen & Tajvidi 2000) and losses on bonds and loans (Chava et al. 2008). We use this exceedances over threshold method, since it uses

---

<sup>5</sup>violating the i.i.d. assumption

data more efficiently and we want to stay in line with previous and comparable researches. The distribution which matches the exceedances over threshold is the GPD.

We use two different ways to model the data jointly: copulas and a model based on MEVT. A copula is a multivariate cumulative density function that links marginal cumulative density functions to a joint cumulative density function. This is a very famous result from Sklar (1959) and proved by Schweizer & Sklar (1983). Nelsen (1999) described a tremendous amount of copula theory with many examples of bivariate copulas. Bedford & Cooke (2002) introduced a completely new way of creating complex multivariate dependence structures based on Joe (1996) using parallels to hierarchical modeling (Green & Richardson 2003). In Aas et al. (2009) this methodology is used in order to decompose a multivariate density into a cascade of pair-copulas. In other words, these pair-copulas will be used as simple building blocks to construct higher dimensional dependence structures. We refer to these compositions as vines. A vine however is not a unique decomposition but a set of decompositions. Aas et al. (2009) only considered two different structures, the canonical vine (C-Vine) and the D-Vine. Aas et al. (2009) compared the C- and D-Vine structure with pair-copulas to the widely used multidimensional t-copula. The C- and D-Vine produced a significantly better in-sample fit in terms of likelihood than the t-copula. As mentioned before there are numerous ways to break down a high dimensional system into pair-copulas and therefore numerous different vine structures. Bedford & Cooke (2002) introduced a subset of these vines, the regular vine (R-Vine). Kurowicka (2010) proposed a new method to select a structure based on the largest partial correlation. Dißmann (2010) used rank correlation to select a vine structure. Dißmann (2010) showed that the use of his sequential algorithm on financial datasets is quite often better than using the t-copula. This result was confirmed by Aas (2013), who used the sequential algorithm of Dißmann (2010) on 19 financial variables that constitute the financial market portfolio of the DNB (Den Norske Bank). To the best of our knowledge these methods have not yet been applied in the field of risk management and VaR estimation. The use of copulas in the field of market risk management is however quite common. One of the pioneers of applying copula theory to VaR estimation are Nyström & Skoglund (2002). They use the two-step method from McNeil & Frey (2000) to model the marginals, which are then jointly modeled by a t-copula in the second step. Many variations of this approach have been studied. For example Hsu et al. (2012) used a similar approach with different copulas to show that for emerging markets the use of a Clayton copula is beneficial. Other variations to the method of Nyström & Skoglund (2002) can be found in Liu & Luger (2009), Berg (2009) and Aas et al. (2009).

Besides vine copulas, we use a method based on MEVT to model our data jointly. Throughout the history MEVT is used in multiple disciplines, from an assessment of Heffernan & Tawn (2004) on environmental impact to financial risk management (Embrechts et al. 2003) and Internet traffic modeling. While extreme value theory has existed for quite a long time, constructing a multivariate variant is recognized as a difficult task. The existing methods mostly rely on the use of copulas (Hilal et al. 2011). Heffernan & Tawn (2004) came up with a practical method to model EVT in multiple dimensions. They also overcame the limiting argument in which all variables become large at the same rate. They impose a conditional approach to model extremal dependence. Existing EVT methods assume asymptotic dependence once the assumption of independence does not hold anymore. This leads to an overestimation in the case of joint extreme events (Ledford & Tawn 1996). Hilal et al. (2011) uses the model of Heffernan & Tawn (2004) on bivariate financial data by creating a hedge portfolio. Hilal et al. (2014) extent their earlier work to accommodate more than two dimensions. To the best of our knowledge this methodology has not been applied in the field of VaR estimation.

### 3 Data

Since we are interested in modeling dependencies, we choose our portfolio in such a way that there is a fair amount of correlation present. We choose to construct a portfolio containing several different major world indices <sup>6</sup>. Other researches in this domain (Hilal et al. (2014), Dißmann (2010)) also chose a portfolio containing major world indices or a portfolio containing exchange rates. The data consists of end of the day closing prices sampled on daily basis.

---

<sup>6</sup>AEX, CAC40, DAX30, Hang Seng, NIKKEI225, SMI and S&P500

The dataset will be discussed in more detail in Section 3.2. Firstly we focus on the data transformation, which is reported in Section 3.1.

### 3.1 Data Transformation

Once the data is collected, we transform the data to log returns. Let  $P_t$  be the price of a certain asset at time  $t$ , then the log transformation can be described as follows:

$$R_t = \log P_t - \log P_{t-1}, \quad (1)$$

where  $R_t$  denotes the log return of a certain asset at time  $t$ . The log transformation is a standard in finance and is also being used in comparable researches. During the weekends we assume that there is no economic activity. Therefore we calculate the over weekend log return as the  $\log P$  at Monday minus the  $\log P$  at Friday. In the case of missing data both the return on that particular day and the return the next day is removed from the entire dataset. Missing data may occur when a certain market is closed. This means that there is practically no economic activity for that certain market, however other financial markets may be open and active. Our data is sampled from different markets and if we would delete only the return on that particular day, we assume that all the markets are inactive. Therefore we choose to remove the return on the particular day and the return on the next day for all the indices in our dataset.

### 3.2 Indices Dataset

The indices dataset is constructed using seven different indices, namely: AEX, CAC40, DAX30, HangSeng, NIKKEI225, SMI and S&P500 <sup>7</sup>. We choose to use this combination of indices because the geographical locations of the indices are spread across different continents and the indices are mutual differently financially integrated. The data is initially quoted in the local currency. Therefore the price movement over time is also influenced by the movement of the exchange rate market. In order to isolate the return on the indices, the exchange rate effects are filtered out by transforming the local currency to USD using the closing foreign exchange rate of this particular currency against the USD.

The indices dataset consists of daily data starting from the 5<sup>th</sup> of January 1999 until the 17<sup>th</sup> of November 2017. After having filtered out the missing data, 3719 observations are left. The price movements over time of several indices are given in Figures 1 and 2. In Figure 1 these movements are shown for the CAC40, DAX30 and SMI. In Figure 2 these movements are shown for the AEX and S&P500 <sup>8</sup>. Both figures indicate a dependence in price movement. Shocks in prices occur quite often simultaneously, which can especially be observed during the Financial Crisis in 2008.

---

<sup>7</sup>Data from Yahoo! Finance

<sup>8</sup>We choose to use different plots for scaling issues

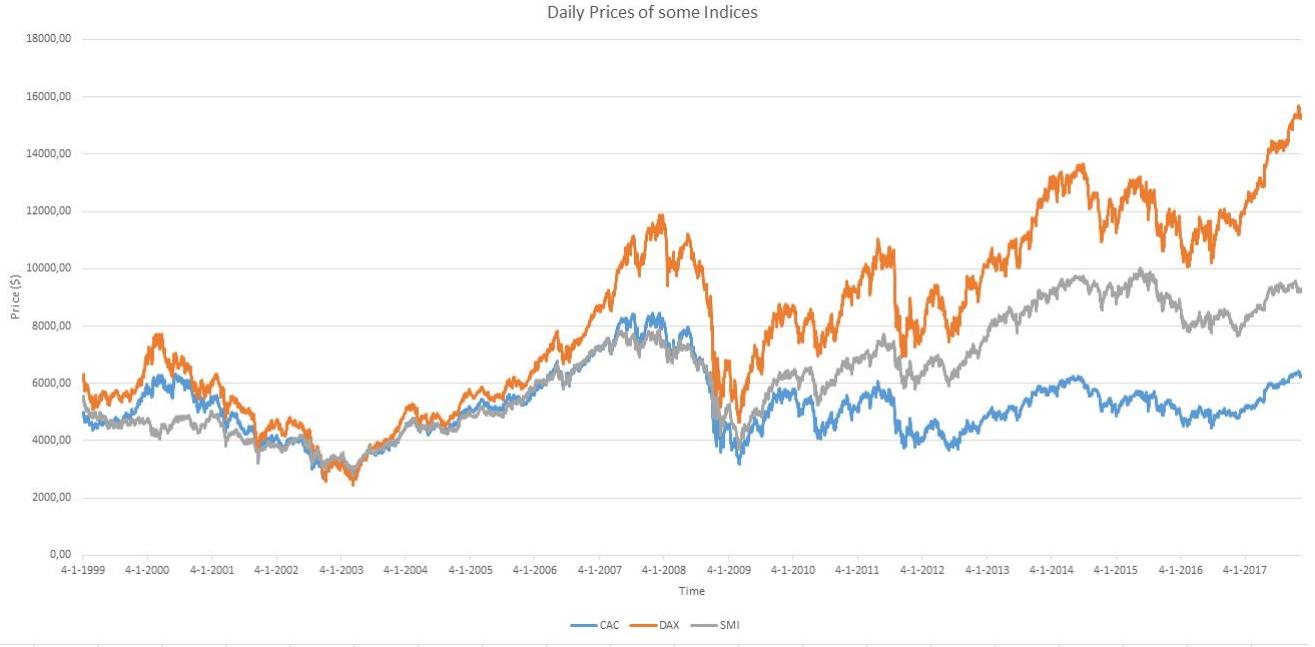


Figure 1: Daily prices of the CAC40, DAX and the SMI over a period of 18 years (5-1-1999 until 17-11-2017)

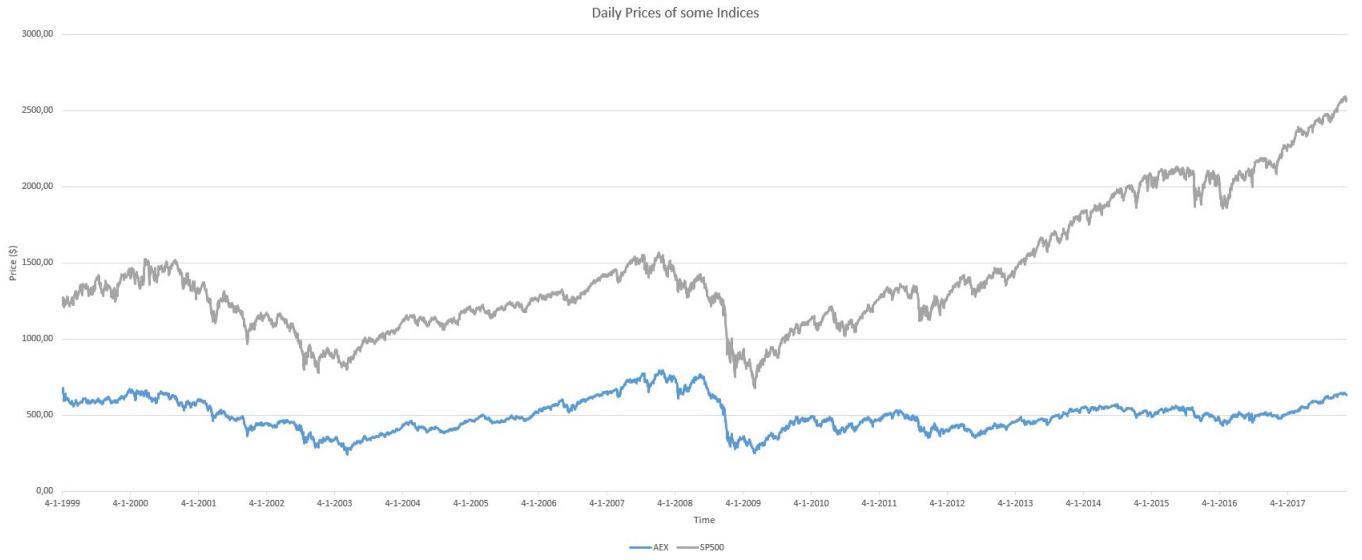


Figure 2: Daily prices of the AEX and the S&P500 over a period of 18 years (5-1-1999 until 17-11-2017)

In Table 1 some descriptive statistics are reported. We observe that the mean is very close to zero for every single index. The standard deviations, maxima and minima of the different indices are quite comparable. High kurtosis indicate the presence of fat tails. The negative skewness indicates that the distributions are skewed to the left. These results are in line with the stylized facts from McNeil et al. (2005).

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
Mean	-0,0001	0,0001	0,0002	0,0001	0,0000	0,0001	0,0001
Maximum	0,1166	0,1179	0,1093	0,1340	0,0950	0,0809	0,1025
Minimum	-0,1184	-0,1173	-0,0959	-0,1359	-0,1111	-0,0773	-0,0947
Skewness	-0,4031	-0,2150	-0,2554	-0,0024	-0,3182	-0,2657	-0,4275
Standard Deviation	0,0153	0,0158	0,0160	0,0148	0,0150	0,0120	0,0121
Kurtosis	9,8226	8,9990	7,1618	11,2466	6,6085	7,4490	10,7397

Table 1: Descriptive statistics of the daily log returns of some major indices: AEX, CAC 40, DAX, Hang Seng, NIKKEI 225, SMI and S&P500 for a period from 5-1-1999 until 17-11-2017.

Since we are interested in dependences among different indices, we construct a figure, which can be found in the Appendix C in Figure 8, containing the Kendall's  $\tau$ <sup>9</sup> in the upper triangle and the scatter plot in the lower triangle for all the different pairs. This figure indicates the presence of strong dependences between the European indices. Furthermore a strong dependence can be observed between the two Asian indices. The S&P500 dataset seems to be more strongly linked to the European indices than to the Asian indices.

## 4 Methodology

To determine the VaR or the ES of a portfolio we follow two consecutive steps. Firstly we model the marginals by our own extension of the two-step method from McNeil & Frey (2000). Once we estimate both steps, we transform our data to the standard uniform scale. Secondly we use this transformed data to estimate the joint distribution by either Vine Copulas or a distribution based on MEVT. This section therefore is split up into two different parts: marginal modeling and joint modeling. However before stepping into these parts, we believe it is important to put risk and dependency into a mathematical context, which can be found in Sections 4.1 and 4.2. In Sections 4.3 and 4.4 both steps of the two-step method (ARMA-MS-GARCH-GJR filter and Semi-Parametric using EVT respectively) are discussed. We will emphasize on the joint modeling based on (Vine) copulas in Section 4.5 and on MEVT in section 4.7. In order to evaluate the VaR and ES estimation performance of our dependence model we use backtests. Section 4.8 we describe several different backtests.

### 4.1 Risk Measures

Risk Measures allow us to quantify risk. Quantifying risk is important in order to analyze and manage it later on. Artzner et al. (1999) defined requirements for a so-called coherent risk measure. Artzner et al. (1999) basically defined four properties that a appropriate risk measures should have. Let  $X$  be a random variable with a return distribution  $F$ . Consider a set of random variables  $\mathcal{M}$  and  $\rho$  a risk measure such that  $\rho : \mathcal{M} \rightarrow \mathbb{R}$ . Then for  $\rho$  to be coherent the following properties need to be satisfied:

1. Monotonicity

For  $X_1, X_2 \in \mathcal{M}$  such that  $X_1 \leq X_2$  it holds that:

$$\rho(X_1) \leq \rho(X_2)$$

This means that when the worth of a object is smaller than the worth of another object, the object with the highest worth is more risky.

2. Translation Invariance

For all  $X \in \mathcal{M}$  and every constant  $A \in \mathbb{R}$  it holds that:

$$\rho(X + A) = \rho(X) - A$$

<sup>9</sup>More information can be found in Section 4.2.1.

This shows that addition of capital reduces risk by the same amount.

### 3. Positive Homogeneity

For all  $X \in \mathcal{M}$  and every constant  $A > 0$  it holds that:

$$\rho(AX) = A\rho(X)$$

This shows that the risk scales together with the size.

### 4. Subadditivity

For  $X_1, X_2 \in \mathcal{M}$  it holds that:

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$$

This means that merging does not create risk. This property is closely related to the benefit of diversification.

In the next section we elaborate on two different risk measures, that are widely used in practice.

#### 4.1.1 Value-at-Risk (VaR)

Value-at-Risk or VaR is the most widely used risk measure in the financial industry and plays a significant role in the Basel agreements. Given a certain confidence level, say  $\alpha$ , the  $\text{VaR}_\alpha$  corresponds to the loss which has a probability  $1 - \alpha$  of being exceeded by a larger loss. When a loss (L) (profit (P)) distribution is considered, the VaRs focus lies on the upper (lower) quantile  $1 - \alpha$  ( $\alpha$ ):  $\text{VaR}_{1-\alpha} = \inf\{l \in \mathbb{R} : P(L > l) \leq \alpha\}$  (quantile  $\alpha$ :  $\text{VaR}_\alpha = \sup\{p \in \mathbb{R} : P(P < p) \leq \alpha\}$ ). Artzner et al. (1999) criticized the VaR for being not coherent. Therefore, Artzner et al. (1999) introduced a coherent alternative of the VaR, the Expected Shortfall (ES) or the Conditional VaR (CVaR).

#### 4.1.2 Expected Shortfall (ES)

Expected Shortfall or ES is a risk measure that is very similar to the VaR. Where VaR violates the Subadditivity property, ES does not. Expected shortfall given a certain confidence level  $\alpha$  is equal to:  $ES_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u du$ . The integral in the expression of the ES presented above, evaluates the total volume of the losses in the tail, or the  $\alpha \times 100\%$  greatest losses. The ES is then obtained by dividing this volume by the relative size of the tail,  $1 - \alpha$ . The difference between the VaR and the ES is that the VaR tells us nothing about the size of the potential loss when the threshold is exceeded, unlike the ES. The ES focuses on the expected loss when a certain loss is exceeded. The ES is more sensitive to the shape of the tail.

## 4.2 Dependence Measures

In this section we describe dependence in a mathematical context. We think this is important, since dependencies play a key role in our research. A dependence measure basically quantifies the level of dependence between two or more variables. A well-known dependence measure is Pearson's  $\rho$ . Pearson's  $\rho$  is a linear measure between two random variables that is often referred to as correlation. As we work with the two-step method from McNeil & Frey (2000) and with copulas, where data transformation plays a significant role, it is important to use a risk measure that is invariant under marginal transformations. Pearson's  $\rho$  correlation is not invariant under marginal transformations. We narrow our scope to two potential candidates, which are invariant under marginal transformations: Spearman's  $\rho$ , which measures the rank correlation and Kendall's  $\tau$ , which measures the level of concordance. Both the Spearman's  $\rho$  and the Kendall's  $\tau$  are linked to copulas. Since both measures are closely related and we want to stay as much as possible in line with previous researches in this domain (Dissmann 2010), we use Kendall's  $\tau$  as dependence measure. We elaborate on Kendall's  $\tau$  in section 4.2.1.

It is widely known that the dependence structure in the main body of the distributions can behave quite differently from the dependence structure in the tails of the distributions. To measure the dependencies in these particular regions a tail dependence measure is used. In section 4.2.2 we describe tail dependence.

### 4.2.1 Kendall's $\tau$ ( $\tau$ )

Kendall's  $\tau$  is a measure of concordance, which assesses the ordinal association between two variables, say  $X, Y$ . To explain concordance, let us denote two random vectors in  $\mathbb{R}^2$  by  $(X_1, Y_1)$  and  $(X_2, Y_2)$ . These variables are considered to be concordant if  $(X_1 - X_2)(Y_1 - Y_2) > 0$  and discordant if  $(X_1 - X_2)(Y_1 - Y_2) < 0$ . Now Kendall's  $\tau$  can be expressed as follows:

$$\tau(X, Y) = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0) = \mathbb{E}[\text{sign}((X_1 - X_2)(Y_1 - Y_2))] \quad (2)$$

where  $\tau \in [-1, 1]$ ,  $(X_1, Y_1) \sim F$  and  $(X_2, Y_2) \sim F$  are independent pairs of random variables.  $\text{sign}$  indicates a function that gives minus one as output whenever the input is negative and one otherwise. Kendall's  $\tau$  can be estimated using:

$$\hat{\tau}(X, Y) = \binom{n}{2}^{-1} \sum_{1 \leq t < s \leq n} \text{sign}((X_t - X_s)(Y_s - Y_t)) \quad (3)$$

It can be proven that the Kendall's  $\tau$  is linked to copulas (see Fredricks & Nelsen (2007) for the proof). In the bivariate case the Kendall's  $\tau$  is linked to a bivariate copula with copula variables  $u_1$  and  $u_2$  as follows:

$$\tau = 4 \int_{u_2} \int_{u_1} C(u_1, u_2) dC(u_1, u_2) - 1 \quad (4)$$

This expression is useful when one wants to estimate copula variables (more on that in Section 4.6.3).

### 4.2.2 Tail Dependence

It is well-known that the joint behavior of two variables around the mean can be quite different from the behavior in the tails. In this section we quantify the joint behavior in the tails by introducing upper and lower tail dependence. In the case of upper tail dependence we focus on the right tail of the distribution. If we consider two variables  $X$  and  $Y$ , then the upper tail dependence of these variables can be expressed as the probability that one variable is in the upper  $q^{th}$  quantile, given the fact that the other variable is in this quantile. Now the upper tail ( $\lambda_u$ ) dependence can be expressed as follows:

$$\lambda_u(X, Y) = \lim_{q \rightarrow 1^-} P(Y > F_Y^\leftarrow(q) | X > F_X^\leftarrow(q)) \quad (5)$$

where  $F_X$  and  $F_Y$  are the marginal cumulative density functions of  $X$  and  $Y$  and provided that  $\lambda_u(X, Y) \in (0, 1]$  exists. Where we define  $F^\leftarrow(q) = \inf\{X \in \mathbb{R} : F(X) \geq q\}$ . Using the same reasoning, we can define lower tail dependence as follows:

$$\lambda_l(X, Y) = \lim_{q \rightarrow 0^+} P(Y \leq F_Y^\leftarrow(q) | X \leq F_X^\leftarrow(q)), \quad (6)$$

provided that the limit  $\lambda_l(X, Y) \in (0, 1]$  exists.

When the tail dependence equals zero then the random variables are considered to be asymptotic independent, whereas the tail dependence is larger than zero the random variables are considered to be asymptotically dependent.

## 4.3 GARCH modeling

In this section we pay attention to the first step in the two-step approach from McNeil & Frey (2000). This step contain the data filtering by using a GARCH model. Instead of using a GARCH model we extent the GARCH by adding an ARMA part to capture possible autocorrelations and a GJR part to capture the leverage-effect. This is discussed Section 4.3.1. Furthermore we choose to govern the data by a non-observed two-state Markov chain, which details are discussed in Section 4.3.2. Finally in order to test the performance of our filter, we provide tests for i.i.d. data and stationarity, i.e. the Ljung Box test and the Augmented Dickey-Fuller test.

### 4.3.1 ARMA-GARCH-GJR

Both the copula and the multivariate extreme value theory require i.i.d. data as input. It has already been pointed out that the financial returns feature some non-i.i.d. characteristics. The ARMA-MS-GARCH-GJR model can be used to filter the returns series (Poon et al. 2004) in order to obtain data that is (close to) i.i.d. data. In this Section we focus on the ARMA-GARCH-GJR and in Section 4.3.2 we elaborate on the Markov-Switching variant of this model.

The ARMA-GARCH-GJR basically consists of three different parts. Firstly the ARMA part, which is able to capture potential autocorrelation in our data. Since it is well-known that financial returns barely show autocorrelation, we expect that the influence of ARMA is rather limited. Secondly there is the GARCH part. This part captures the volatility movement and clustering over time. Thirdly we have the GJR part, which accounts for asymmetry and the leverage effect.

The ARMA-GARCH-GJR was firstly introduced by Glosten et al. (1993). A downside of these kind of models is that a GARCH type requires an assumption about the residual distribution. Because it is widely known that many financial time series have fat tails (leptokurtic tails), a student t distribution is chosen as underlying residual distribution. The degrees of freedom indicate the tail heaviness.

If  $R_t$  represent the financial return at time  $t$ , then the ARMA-GARCH-GJR with respectively  $(i, j)$  ARMA lags,  $(k, l)$  GARCH lags and  $m$  asymmetric leverage parameters, can be described as follows <sup>10</sup>:

$$\begin{aligned} R_t &= \mu_t + \sigma_t \cdot X_t, \\ \mu_t &= \mu + \psi_1 \cdot R_{t-1} + \dots + \psi_i \cdot R_{t-i} + \theta_1 \cdot \epsilon_{t-1} + \dots + \theta_j \cdot \epsilon_{t-j}, \\ \sigma_t^2 &= \omega + \alpha_1 \cdot \epsilon_{t-1}^2 + \dots + \alpha_k \cdot \epsilon_{t-k}^2 + \gamma_1 \cdot I_{t-1} \cdot \epsilon_{t-1}^2 + \dots + \gamma_m \cdot I_{t-m} \cdot \epsilon_{t-p}^2 + \beta_1 \cdot \sigma_{t-1}^2 + \dots + \beta_l \cdot \sigma_{t-l}^2, \\ \epsilon_t &= R_t - \mu_t, \\ X_t &\sim t(\nu). \end{aligned} \quad (7)$$

Where  $I$  is an indicator function which takes on value one if  $\epsilon \leq 0$  and zero otherwise.  $X_t$  is student-t distributed with  $\nu$  degrees of freedom. For stationarity and invertibility of the conditional mean  $\mu_t$ , the following constraints are imposed:  $|\psi_i| < 1$  for  $i = 1, \dots, m$ ,  $|\theta_i| < 1$  for  $i = 1, \dots, n$ ,  $\omega > 0$ ,  $\alpha_i \geq 0$  for  $i = 1, \dots, p$ ,  $\beta_i \geq 0$  for  $i = 1, \dots, q$  and finally  $\sum_i^p \alpha_i + \sum_i^q \beta_i < 1$ .  $\sigma_t$  denotes the conditional variance at time  $t$ ,  $\omega$  the intercept and  $\epsilon_t$  the residuals at time  $t$ .

To estimate and use the model the number of lags,  $i, j, k, l$  and  $m$  need to be calibrated. on one hand we want to avoid overfitting, while on the other hand we want to incorporate as much informative lags as possible. Therefore choosing the number of lags is a balance between overfitting the in-sample and producing an accurate model. Brockwell & Davis (1991) use the information criteria to optimize this balance. They fit the model is fitted using various lags. The number of lags that optimizes the information criterion, is considered to be the most suited. There are two different information criteria that are common in market practice:

- Firstly there is the Alkaike Information Criterion (AIC). The AIC quantifies a balance between the model fit in terms of the log likelihood and the number of parameters used in the model. It can be described as follows:

$$AIC = -2 \log \hat{L} + 2n_p, \quad (8)$$

where  $\log \hat{L}$  represents the log likelihood and  $n_p$  the number of parameters used in the model. It follows that the AIC decreases (increases) when the model fit increases (decreases) or/and the number of parameters in the model decreases (increases). In other words the model with the lowest AIC is considered to optimal.

- Secondly there is the Bayes Information Criterion (BIC). The BIC is very similar to the AIC. The BIC value however is also influenced by the number of observations in the dataset. The BIC quantifies a balance

---

<sup>10</sup>In market practice  $m$  is set equal to  $l$

between the model fit in terms of the log likelihood and the number of parameters **and** number of observations used in the model. The BIC can be described as follows:

$$BIC = -2 \log \hat{L} + n_p \log n, \quad (9)$$

where  $n$  represents the number of observations. Again it follows that the model that leads to the lowest BIC value is preferred.

The ARMA-GARCH-GJR model will be fitted using quasi maximum likelihood estimation (QMLE). Once the number of lags is selected by one of the information criteria, we choose the appropriate number of lags. The model is fitted based on this number of lags. The fitted values will be used to filter the returns, following Poon et al. (2004). This process can be described as follows:

$$X_t = \frac{R_t - \hat{\mu}_t}{\hat{\sigma}_t}, \quad (10)$$

where  $X_t$  is assumed to be approximately i.i.d. over time.  $\hat{\mu}_t$  and  $\hat{\sigma}_t$  represents the fitted conditional mean and variance on time  $t$ , respectively. To test if  $X_t$  is i.i.d. over time we apply the Ljung Box and the Augmented Dickey-Fuller tests, which are described respectively in Sections 4.3.3 and 4.3.4.

### 4.3.2 ARMA-MS-GARCH-GJR

Mikosch & Starica (2004) show that the estimation of GARCH models on samples with displaying structural changes in the unconditional variance, have a bias due to the inflexible nature when GARCH parameters are held fixed. We think it is reasonable to assume that our data contains different regimes over time. After all our dataset contains the Financial Crisis in 2008. Therefore we believe that it might be advantageous to make the ARMA-GARCH-GJR model more flexible by allowing parameters to differ across different regimes. The regimes are however unobserved and can be determined by applying the Hamilton filter (Bauwens et al. 1989). We can use the Markov-Switching GARCH model from Bauwens et al. (2010) where the conditional mean and variance switch in time from one GARCH process into another governed by a non-visible Markov-Chain. The ARMA-MS-GARCH-GJR is quite similar to the ARMA-GARCH-GRJ model described in equation 7. The conditional volatility process differs among different regimes and can be described as follows:

$$\begin{aligned} \sigma_t^2 = & (\omega_1 + \alpha_{11} \cdot \epsilon_{t-1}^2 + \dots + \alpha_{1k} \cdot \epsilon_{t-k}^2 + \gamma_{11} \cdot I_{t-1} \cdot \epsilon_{t-1}^2 + \dots + \gamma_{1m} \cdot I_{t-m} \cdot \epsilon_{t-p}^2 + \beta_{11} \cdot \sigma_{t-1}^2 + \dots + \beta_{1l} \cdot \sigma_{t-l}^2) \times I\{S_t = 1\} \\ & + (\omega_2 + \alpha_{21} \cdot \epsilon_{t-1}^2 + \dots + \alpha_{2k} \cdot \epsilon_{t-k}^2 + \gamma_{21} \cdot I_{t-1} \cdot \epsilon_{t-1}^2 + \dots + \gamma_{2m} \cdot I_{t-m} \cdot \epsilon_{t-p}^2 + \beta_{21} \cdot \sigma_{t-1}^2 + \dots + \beta_{2l} \cdot \sigma_{t-l}^2) \times I\{S_t = 2\} \\ & \vdots \\ & + (\omega_q + \alpha_{q1} \cdot \epsilon_{t-1}^2 + \dots + \alpha_{qk} \cdot \epsilon_{t-k}^2 + \gamma_{q1} \cdot I_{t-1} \cdot \epsilon_{t-1}^2 + \dots + \gamma_{qm} \cdot I_{t-m} \cdot \epsilon_{t-p}^2 + \beta_{q1} \cdot \sigma_{t-1}^2 + \dots + \beta_{ql} \cdot \sigma_{t-l}^2) \times I\{S_t = q\} \end{aligned} \quad (11)$$

Where  $S_t$  represent the state and  $I\{S_t = x\}$  an indicator function which equals one if the state is equal to  $x$  and zero otherwise. We assume that there are  $q$  different states. If we would consider two different Markov states, then  $S_t \in \{1, 2\} \quad \forall t$ . The ARMA-MS-GARCH-GJR model can be estimated by using QML. We choose to the student-t distribution as underlying distribution for the innovations. To limit the computational time we only assume that there are two different states. The likelihood function can be described as follows:

$$\mathcal{L}(\Psi | I_T) = \prod_{t=1}^T \left[ \sum_{i=1}^2 \sum_{j=1}^2 p_{ij} Pr(S_{t-1} = i | \Psi, I_{t-1}) f(X_t | S_t = j, \Psi, I_{t-1}) \right] \quad (12)$$

where  $\mathcal{L}$  represent the likelihood function and

$$f(X_t|S_t = j, \Psi, I_{t-1}) = f\left(\frac{R_t - \mu_t}{\sigma_t} | S_t = j, \Psi, I_{t-1}\right) \quad (13)$$

$$X_t \sim t(\nu)$$

The  $Pr(S_{t-1} = i | \Psi, I_{t-1})$  represent the filtered probability of state  $i$  when using the Hamilton filter.  $p_{ij}$  represents the  $(i, j)^{th}$  element of the transition matrix  $P$ <sup>11</sup>.  $\Psi$  corresponds to the vector containing the model parameters  $(\phi_1 \dots \phi_i, \omega_1, \alpha_{11}, \dots, \alpha_{1k}, \beta_{11}, \dots, \beta_{1l}, \gamma_{11}, \dots, \gamma_{1m}, \omega_2, \alpha_{21}, \dots, \alpha_{2k}, \beta_{21}, \dots, \beta_{2l}, \gamma_{21}, \dots, \gamma_{2m})$

#### 4.3.3 Ljung Box test

The i.i.d. assumption can be tested using the Ljung Box test. This is an independence test originally founded by Ljung & Box (1978). There are several other ways to test for independence. We choose the Ljung Box test, because it is commonly used in practice. Under the null hypothesis we assume that the data is randomly distributed (i.e. correlations are zero). The alternative hypothesis is that the data are not randomly distributed and that they exhibit serial correlation. The Ljung Box test statistic  $Q$  is defined as:

$$Q = n(n+2) \sum_{i=1}^k \frac{\hat{\rho}_i^2}{n-i} \quad (14)$$

where the sample size is denoted by  $n$ , the squared empirical autocorrelation at lag  $i$  by  $\hat{\rho}_i^2$  and the number of lags that are included in the test by  $k$ . Under the null hypothesis it follows that  $Q$  behaves as a chi-squared distribution with  $k$  degrees of freedom ( $\chi_k^2$ ). In other words, given a significance level  $\alpha$ , the null hypothesis can be rejected if:

$$Q > \chi_{1-\alpha, k}^2$$

where  $\chi_{1-\alpha, k}^2$  is the  $1 - \alpha$  quantile of the  $\chi_k^2$  distribution with  $k$  degrees of freedom.

#### 4.3.4 Augmented Dickey-Fuller test

The Dickey-Fuller test checks for the presence of a unit root in an autoregressive model. Under the null hypothesis we assume that there is a unit root present. The test equation can be formulated as follows:

$$\Delta Y_t = \theta_0 + \gamma t + \rho Y_{t-1} + \delta_1 \Delta Y_{t-1} \dots + \delta_{p-1} \Delta Y_{t-p+1} + \epsilon_t \quad (15)$$

where the coefficients are estimated by using OLS and the t-statistic of  $\hat{\rho}$  is used to test for a unit root. The null hypothesis,  $\rho = 0$ , is tested against the alternative hypothesis,  $\rho < 0$ .

### 4.4 Semi-Parametric modeling and Extreme Value Theory (EVT)

Once the marginals are filtered using the ARMA-MS-GARCH-GJR model, we use semi-parametric modeling to model the filtered returns, which is the second step in the two-step approach of McNeil & Frey (2000). Semi-parametric modeling means that we model the distribution by both a parametric and a non-parametric part. We model the tails of the distribution by using a distribution based on extreme value theory (GPD). The main body of the data however, will be modeled non-parametrically by the empirical distribution.

Using parametric models to determine the tail behavior of distributions is beneficial because of several reason. Firstly it is worth mentioning that VaR or ES estimations focus primarily on the tail regions. However these regions are characterized by a data scarcity. Parametric models for the tails are therefore quite informative. Moreover these parametric models allow for inter- and extrapolation. This is a huge benefit once we want to

---

<sup>11</sup>For details about the Hamilton filter see Appendix A

transform the simulated data on standard uniform scale to simulated returns.

Basically there are two well known approaches for putting extreme value theory into practice. The first approach is the block maxima approach. The disadvantage of this approach is that the data is used less efficiently compared to the second approach, the peaks of threshold approach (POT approach). Following McNeil & Frey (2000) and Hilal et al. (2011) we prefer to model the tails using GPD which is in line with the POT approach. Consider random variable  $X$ , then the GPD ( $G$ ) is given by:

$$G_{\xi, \beta}(x) = \begin{cases} 1 - (1 + \xi x / \beta)^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp(-x / \beta), & \xi = 0 \end{cases} \quad (16)$$

where  $\beta > 0$ , and  $x \geq 0$  when  $\xi \geq 0$  and  $0 \leq x \leq -\beta/\xi$  when  $\xi < 0$ . The parameter  $\xi$  refers to the shape, whereas the parameter  $\beta$  refers to the scale. We can now describe the entire marginal distribution (body and tails)  $F_x$ , for each marginal separately as follows <sup>12</sup>:

$$\widehat{F}_X(x) = \begin{cases} \widetilde{F}_X(u_X^-)[1 + \xi^-(u_X^- - x) / \beta^-]^{-1/\xi^-} & \text{if } x < u_X^-, \\ \widetilde{F}_X(x) & \text{if } x \in [u_X^-, u_X^+], \\ 1 - \widetilde{F}_X(u_X^+)[1 + \xi^+(x - u_X^+) / \beta^+]^{-1/\xi^+} & \text{if } x > u_X^+ \end{cases} \quad (17)$$

Where  $\widetilde{F}_X(x) = \sum_{t=1}^n \frac{1}{n} \mathbb{I}(X_t \leq x)$  is the empirical distribution function of variable  $X$  and  $\widetilde{F}_X(u_X^+) = 1 - \widetilde{F}_X(u_X^+)$  is the survivor function.  $u_X^+$  and  $u_X^-$  stand for the upper and lower threshold respectively, while  $\xi^+$ ,  $\beta^+$ ,  $\xi^-$  and  $\beta^-$  stand for the positive shape and scale parameters and the negative shape and scale respectively. The shape and the scale can be estimated in different ways. We choose to estimate these parameters by using maximum likelihood estimation.

However we did not yet define the thresholds for the marginals,  $u_X^+$  and  $u_X^-$ . The choice of a proper threshold is a balance between bias and variance. Allowing lots of observations in the tail will lead to smaller variance and larger bias, while allowing fewer observations will result to larger variance and smaller bias. There is no straightforward approach on how to determine these thresholds. In practice, eyeball tests can be used to determine these thresholds properly (McNeil et al. 2005).

The first method is based on the one of the characteristics of the GPD. A linear relationship between the mean excess over a certain threshold and the particular threshold itself can be observed when the threshold is chosen properly. The mean excess function over threshold  $u$  ( $e(u)$ ) can be defined as follows:

$$e(u) = \frac{\beta + \xi u}{1 - \xi},$$

where  $0 \leq u < \infty$  if  $0 \leq \xi < 1$  and  $0 \leq u \leq -\beta/\xi$  if  $\xi < 0$ .  $e(u)$  can empirically be estimated by:

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u) I_{\{X_i > u\}}}{\sum_{i=1}^n I_{\{X_i > u\}}} \quad (18)$$

When the thresholds ( $u$ ) are plotted against their corresponding empirically estimated mean excess ( $e_n$ ) at a certain point linear relationship can be found. A threshold within this particular region can be seen as a properly chosen one. This method requires however an eyeball test and can therefore be seen as arbitrary test McNeil et al. (2005).

Another way to estimate the threshold is by applying an eyeball test to hill plots. If we order the observations from small to big, such that  $X_{(1)} > X_{(2)} > \dots > X_{(n)}$ , then we can express the Hill as (McNeil et al. 2005):

$$\hat{\alpha}_{k,n} = \left( \frac{1}{k} \sum_{j=1}^{(k-1)} \log X_{(j)} - \log X_{(k)} \right), \quad 2 \leq k \leq n \quad (19)$$

---

<sup>12</sup>the proof can be found in the Appendix A.2

where  $\hat{\alpha}_{k,n}$  is the empirical estimated tail index when there are  $k$  exceedances over the threshold. The tail index is denoted by  $\alpha = 1/\xi$ . In practice different values of  $k$  are plotted against their corresponding  $\hat{\alpha}_{k,n}$ . The first stable regions indicates a proper chosen number threshold.

Since we want to backtest our models the marginal and the joint modeling needs to be repeated over and over again. Applying eye-ball tests to every single hill plot is absolutely infeasible. Therefore we choose to apply the automated eye-ball method from Danielsson et al. (2016). This automated approach is based on the acquiring a stable region within Hill-plot. Let  $k_{eye}^*$  be the by the automated eye-ball test selection number of exceedances. Then  $k_{eye}^*$  can be estimated using:

$$k_{eye}^* = \min \left\{ k \in 2, \dots, (n-w) \left| h < \frac{1}{w} \sum_{i=1}^w I[\hat{\alpha}_{k+i,n} < \hat{\alpha}_k \pm \epsilon] \right. \right\}, \quad (20)$$

where  $\hat{\alpha}_k$  is the Hill estimation of the tail index and the values  $w$ ,  $h$  and  $\epsilon$  need to be chosen specifically to the data.  $w$  is the size of the moving window in which the stability is checked,  $\epsilon$  gives the range in which  $\hat{\alpha}$  is considered to be stable and  $h$  is the percentage of estimates which should be within this range. According to McNeil et al. (2005) it is misleading to choose a large  $k$  when looking at Hill-plots. McNeil et al. (2005) suggest that a low number of exceedances is often the best choice.

## 4.5 Copulas

Using the two step approach previously explained, the data is transformed into observations on a standard uniform scale that is (close to) i.i.d.. This is required when working with copulas. Copulas can be used to act as a bridge between marginal distributions and joint distributions. The name "copula" comes from Latin, which means "link". In other words, copulas are used to link several marginal distributions to a joint distribution. There are several different ways to link (i.e. different copulas) the marginals. Since copulas reveal dependence on the standard uniform scale, copulas can help us to understand dependence structures at a deeper level. In this section basics about copulas are described along with several different copula families.

### 4.5.1 The Concept of Copulas

In this section the basic concepts of copulas are described. Copulas are defined as follows:

#### Definition

A n-dimensional copula  $C$  is a joint distribution on  $[0, 1]^n$ ,  $C : [0, 1]^n \rightarrow [0, 1]$  with univariate uniform distributed marginals.

Let  $X = (X_1, \dots, X_n)$  be a n-dimensional random variable, with joint probability density function  $f(x_1, \dots, x_n)$  and joint cumulative distribution function  $F(x_1, \dots, x_n)$ . Let  $F_1(x_1), \dots, F_n(x_n)$  be the corresponding strictly increasing continuous marginal cumulative distribution functions of  $X_1, \dots, X_n$ . Let  $f_1(x_1), \dots, f_n(x_n)$  be the corresponding marginal probability density functions.

Sklar's theorem (Sklar 1959) states that a joint distribution can be expressed as a set of marginal distributions and a copula:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \quad \text{where } C \text{ is the copula of } F(x_1, \dots, x_n). \quad (21)$$

By taking the derivative of the copula stated above the joint probability density function can be retrieved. It follows that:

$$c(F_1(x_1), \dots, F_n(x_n)) = \frac{\partial^n C(F_1(x_1), \dots, F_n(x_n))}{\partial F_1(x_1) \cdots \partial F_n(x_n)} \stackrel{21}{=} \frac{\partial^n F(x_1, \dots, x_n)}{\partial F_1(x_1) \cdots \partial F_n(x_n)} = \frac{f(x_1, \dots, x_n)}{\prod_{i=1}^n f_i(x_i)} \quad (22)$$

Using this equation we can derive the joint probability density function:

$$f(x_1, \dots, x_n) = \left( \prod_{i=1}^n f_i(x_i) \right) c_{12\dots n}(F_1(x_1), \dots, F_n(x_n)) \quad (23)$$

This is a useful result, since we use the joint probability density function in the maximum likelihood function to estimate the copula parameter(s).

#### 4.5.2 Copula Families

In this section we describe different famous copula families. We have seen that a copula is capable of linking marginals to a joint distribution. Copulas are capable of linking marginals to a joint distribution in many ways. Different copula families represent different dependence structures. Generally speaking, there are two different classes of copula families. Firstly there are elliptical copulas and secondly there are Archimedean copulas. Throughout the history many different other copulas have been developed, including for example the class of extreme value copulas. In order to avoid additional complexities we choose to use some well-known elliptical and Archimedean copulas.

#### Elliptical Copula

In this section we describe two different elliptical copulas, the Gaussian copula and the  $t$  copula. Elliptical copulas and especially  $t$  copulas are widely used in finance (Bouyé et al. 2000). In Table 2 the copula formula of the Gaussian and  $t$  copula are presented. Furthermore the corresponding rank correlation is shown (in this case Kendall's  $\tau$ ) along with the tail dependence (McNeil et al. 2005).

Name	Formula	Kendall's tau	Tail Dependence
Gaussian	$C_\rho = \Phi(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$	$\tau = \frac{2}{\pi} \arcsin \rho$	No tail dependence
$t$	$C_{\rho,v} = t_v(t_v^{-1}(u_1), t_v^{-1}(u_2))$	$\tau = \frac{2}{\pi} \arcsin \rho$	$2t_{v+1} \left( -\sqrt{\frac{(v+1)(1-\rho)}{1+\rho}} \right)$

Table 2: *Two elliptical copula: their copula formulas, their relationship between the Kendall's tau and the tail dependence.  $\rho$  stands for the Pearson's correlation coefficient (linear correlation),  $\Phi$  for the cumulative standard normal distribution,  $\Phi^{-1}$  for the inverse standard normal cumulative distribution function and  $t_v$  a Student- $t$  cumulative distribution function with  $v$  degrees of freedom.*

#### Archimedean Copula

Besides the elliptical copulas, we use Archimedean copulas. Archimedean copulas are often used in practice as well and have been studied extensively over time. They have proven to be useful for modeling portfolio risk. Archimedean copulas refer to the class of copulas that can be expressed as follows: (McNeil et al. 2005, page. 220)

$$C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2))$$

where  $\psi$  is a generator function, which is strictly decreasing and convex,  $\psi: [0, 1] \rightarrow [0, \infty]$ , such that  $\psi(1) = 0$ . It is known as a strict generator if  $\psi(0) = 1$ . There are many different generator function used over time, leading to different copulas. The following Table 4 represents some basic properties of most prominent bivariate Archimedean Copulas (McNeil et al. 2005, page. 222).

Name	Generator $\psi_\theta(t)$	Kendall's $\tau$	Parameters	Tail Dependence $(\lambda_L, \lambda_u)$
Clayton	$\frac{1}{\theta}(t^{-\theta} - 1)$	$\frac{\theta}{\theta+2}$	$\theta \geq 0$	$(2^{-1/\theta}, 0)$
Gumbel	$(-\ln(t))^\theta$	$1 - \frac{1}{\theta}$	$\theta \geq 1$	$(0, 2 - 2^{1/\theta})$
Frank	$-\ln\left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}\right)$	$1 - \frac{4}{\theta} + 4\frac{D_1(\theta)}{\theta}$	$\theta \in \mathbb{R}$	$(0, 0)$
Joe	$-\ln(1 - (1-t)^\theta)$	$1 + \frac{1}{\theta^2} \int_0^1 (t \ln(t)) (1-t)^{2(1-\theta)\theta} dt$	$\theta \geq 1$	$(0, 2 - 2^{1/\theta})$

Table 3: This table represents the generator functions of some bivariate Archimedean copula along with the corresponding Kendall's  $\tau$  and the upper and lower tail dependence, respectively  $\lambda_U$  and  $\lambda_L$ .  $D_1(\theta)$  represents the Debye function  $D_1(\theta) = \theta^{-1} \int_0^\theta t/(\exp(t) - 1) dt$ .

## BB Copula

Joe (1997) provided several other Archimedean bivariate copulas with two copula parameters. We refer to these copulas as the BB Copulas. We use four different BB copulas in this research: BB1, BB6, BB7 and BB8. The BB1 copula can for example be seen as a generalization of the Gumbel and Clayton copulas (Joe 1997).

Name	Generator $\psi_\theta(t)$	Kendall's $\tau$	Parameters	Tail Dependence $(\lambda_L, \lambda_u)$
BB1	$(t^{-\theta} - 1)^\delta$	$1 - \frac{2}{\delta(\theta+2)}$	$\theta > 0, \delta \geq 1$	$(2^{-1/(\theta\delta)}, 2^{-1/\delta})$
BB6	$(-\log(1 - [1-t]^\theta))^\delta$	$1 + \frac{4}{\delta\theta} \int_0^1 (-\log(1 - [1-t]^\theta)(1-t)(1 - (1-t)^{-\theta})) dt$	$\theta \geq 1, \delta \geq 1$	$(0, 2 - 2^{-1/(\delta\theta)})$
BB7	$(1 - (1-t)^\theta)^{-\delta} - 1$	$1 + \frac{4}{\delta\theta} \int_0^1 (-(1 - (1-t)^\theta)^{\delta+1} \frac{(1 - (1-t)^\theta)^{-\delta} - 1}{(1-t)^{\theta-1}}) dt$	$\theta \geq 1, \delta > 0$	$(2^{-1/\delta}, 2 - 2^{1/\theta})$
BB8	$-\log(\frac{1 - (1-\delta t)^\theta}{1 - (1-\delta)^\theta})$	$1 + \frac{4}{\delta\theta} \int_0^1 (-\log(\frac{1 - (1-t\delta)^\theta - 1}{(1-t\delta)^{\theta-1}})(1-t\delta)(1 - (1-t\delta)^{-\theta})) dt$	$\theta \geq 1, \delta \in [0, 1]$	$(0, 0)$

Table 4: This table represents the generator functions of some BB Archimedean copula along with the corresponding Kendall's  $\tau$  and the upper and lower tail dependence,  $\lambda_U$  and  $\lambda_L$  respectively.

### 4.5.3 Copula Estimation

Copula parameters can be estimated in several different ways. In our research we limit ourselves to using rank correlation and maximum likelihood to estimate copula parameters.

- Rank correlations can be used to calibrate copulas to empirical data. There are two main varieties of rank correlation, Kendall's  $\tau$  and Spearman's  $\rho$ . It can be proven that the following relationship exist:

$$\rho_\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1 \quad (24)$$

The Kendall's  $\tau$  and Spearman's  $\rho$  can be estimated empirically. In the tables 2-4 this expression have been evaluated for the different copulas. Using these empirical results along with these expression, the copula variable can be estimated.

- Using the probability density function of the copulas, maximum likelihood estimation can be used in order to estimate the copula parameters. The probability density function of any copula can be described as in Equation 23.

In this research we use maximum likelihood to estimate the copula parameters. The advantage of using maximum likelihood is that it easily accommodates estimation of copulas with two or more variables.

### 4.5.4 Copula Selection

As mentioned before there are many different bivariate copulas. In this section our focus lies on the determining an appropriate copula for the data. Looking at scatter plots could give us some useful insights which copula might be appropriate. Presence of for example upper tail dependence would motivate the choice of the Gumbel copula, while a copula without upper tails dependence, like the Clayton copula seems to be a bad choice. However we want to motivate our choice of copula as objective as possible. Furthermore analyzing scatter plots for many different pairs over different time periods is from practical point of view not possible. We use AIC/BIC value as

criterion to determine the copula type. This means that we fit all possible copula candidates to the data. The fit in terms of likelihood is used to assess the AIC or the BIC value. Using the lowest AIC or BIC value is in favor. This selection method is used in many comparable researches, such as Aas et al. (2009), Dißmann (2010) and Schepsmeier & Czado (2015).

#### 4.5.5 Independence Test

Since Vine copulas split a multivariate density up into a cascade of different conditional and unconditional bivariate copulas, many bivariate copulas have to be selected and estimated. We therefore want to limit the model from fitting bivariate copulas on variable pairs that are significantly independent from each other. An independence check will be introduced to check if a complicated copula is required for a particular pair. In the case that a pair is indeed close to independent, we use an independence copula to describe their joint behavior. We use the independence test from Genest & Rémillard (2007). This test checks if the empirical estimation of the kendall's  $\tau$ ,  $\hat{\tau}$  significantly differs from zero. If so, an independence copula is used to describe their relationship. The null hypothesis that the empirical kendall's  $\tau$  equals zero, while the alternative hypothesis is that it is not equal to zero. Genest & Rémillard (2007) show that under the null hypothesis, the test statistic  $\hat{\tau}$  approximately normally distributed with zero mean and a variance of  $\frac{2(2N+5)}{9N(N-1)}$ , where  $N$  is the sample size. In other words, if we are not be able to reject the null hypothesis, then we assume that the independence copula describes the dependence structure well enough.

### 4.6 Pair-Copula Constructions

Throughout the history many different copulas have been developed. Most of these copulas only accommodate a very limited number of dimensions. The construction and the extension to more dimensions of these copulas is recognized as a difficult problem. Furthermore the estimation of many multivariate copulas is characterized by long computational time. Due to these limitations the only feasible copulas, from a practical point of view, are the ones in a very limited number of dimensions, i.e. bivariate copulas. This has resulted to the fact that the class of different bivariate copulas is big, while the number of multivariate ones is rather limited. The Gaussian copula and the t copula are exceptions. They can easily accommodate higher dimensions. Moreover most multivariate copulas that actually can be extended to multiple dimensions are recognized by hard and inflexible dependence conditions. This can be illustrated by for example the t copula, which have symmetric tail dependencies for all of the variable pairs.

Regular Vine copulas provide a solution to these problems. These type of copulas can construct a multivariate density function as a product of bivariate conditional and unconditional copula densities. Vine copulas are based on a decomposition of a multivariate density into a cascade of pair-copulas. In other words, pair-copulas will be used as simple building blocks to construct higher dimensional dependence structures. While the computational time of these copulas is still quite long, these copulas can very easily be extended to higher dimensions. Moreover these copulas do not have the hard and inflexible conditions that other multivariate copulas have.

In this Section we firstly focus on different ways to split a multivariate probability density function into multiple pair-copulas, which can be found in Section 4.6.1. This will form a base from which we introduce R-Vines and graph theory in Section 4.6.2. In Section 4.6.3 we pay attention to R-Vine estimation, while in 4.6.4 we focus on sampling from R-Vines. In Section 4.6.5 the entire procedure of choosing and estimating the R-Vine is shown and we provide a way to compare different R-Vines in Section 4.6.6.

#### 4.6.1 Pair-Copula Constructions (PCC)

In this part we will show how to split up a multivariate distribution into multiple pair-copulas. Consider  $n$  random variables  $\mathbf{X} = (X_1, \dots, X_n)$  with a joint probability density function  $f(x_1, \dots, x_n)$ . It follows that the

joint probability density function can be decomposed as

$$f(x_1, \dots, x_n) = f(x_n) \cdot f(x_{n-1}|x_n) \cdot f(x_{n-2}|x_{n-1}, x_n) \cdot \dots \cdot f(x_1|x_2, \dots, x_n) \quad (25)$$

The decomposition 25 is not unique. If we would relabel the variables, other decompositions can be made. This decomposition plays a central role for the Vine copula theory. The conditional probability density functions (for example:  $f(x_{n-1}|x_n)$ ) can be expressed in a product of copula probability density functions and a unconditional probability density function. For illustrating purposes the full decomposition of multivariate distribution of two and three variables are described below.

1. Consider the variable pair  $(X_1, X_2)$ . Using the result of Equation 22 combining with Bayes rule ( $f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)}$ ) we obtain the following expression:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f(x_2)} = \frac{c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \cdot f_2(x_2)}{f_2(x_2)} = c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \quad (26)$$

2. In the same way we can construct the conditional probability density function for three random variables  $(X_1, X_2, X_3)$ .

$$\begin{aligned} f(x_1|x_2, x_3) &= \frac{f(x_1, x_2|x_3)}{f(x_2|x_3)} = \frac{c_{12|3}(F_{13}(x_1|x_3), F_{23}(x_2|x_3)) \cdot f(x_1|x_3) \cdot f(x_2|x_3)}{f(x_2|x_3)} \\ &= c_{12|3}(F_{13}(x_1|x_3), F_{23}(x_2|x_3)) \cdot f(x_1|x_3) \\ &= c_{12|3}(F_{13}(x_1|x_3), F_{23}(x_2|x_3)) \cdot c_{13}(F_1(x_1), F_3(x_3)) \cdot f_1(x_1) \end{aligned} \quad (27)$$

As mentioned before the decomposition is not unique. A simple variation can be found by relabeling the second and third variable:

$$\begin{aligned} f(x_1|x_2, x_3) &= \frac{f(x_1, x_3|x_2)}{f(x_3|x_2)} = \frac{c_{13|2}(F_{12}(x_1|x_2), F_{23}(x_3|x_2)) \cdot f(x_1|x_2) \cdot f(x_3|x_2)}{f(x_3|x_2)} \\ &= c_{13|2}(F_{12}(x_1|x_2), F_{23}(x_3|x_2)) \cdot f(x_1|x_2) \\ &= c_{13|2}(F_{12}(x_1|x_2), F_{23}(x_3|x_2)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \end{aligned} \quad (28)$$

We follow Stöber et al. (2013) by assuming that all bivariate copulas occurring in the constructions do not depend on the variables that are conditioned on. This means that we assume that  $c_{i,j|k}(\cdot, \cdot|x_k) = c_{i,j|k}(\cdot, \cdot)$ .

Using the results derived in Equations 26-28, we can construct a general expression for multivariate conditional probability density functions using conditional copulas as follows:

$$f(x_i|\mathbf{x}_{-i}) = c_{x_i x_j|\mathbf{x}_{-i,-j}}(F_{i,(-i,-j)}(x_i|\mathbf{x}_{-i,-j}), F_{i,(-i,-j)}(x_j|\mathbf{x}_{-i,-j})) \cdot f(x_i|\mathbf{x}_{-i,-j}) \quad (29)$$

where  $\mathbf{x}_{-i} = \{x_1, x_2, \dots, x_n\} \setminus \{x_i\}$  and  $\mathbf{x}_{-i,-j} = \{x_1, x_2, \dots, x_n\} \setminus \{x_i, x_j\}$  ( $(-i, -j) = \{1, 2, \dots, n\} \setminus \{i, j\}$ ). In other words, a multivariate density function can be described as a non-unique product of pair-copulas. It follows that if the number of dimensions increase, then the number of different decompositions grows exponentially.

In order to describe the three dimensional unconditional probability density function we use the results from the two dimensional probability density function combined with the Bayes rule. Using equation 25 we know that the following expression is valid:

$$f(x_1, x_2, x_3) = f_3(x_3) \cdot f(x_2|x_3) \cdot f(x_1|x_2, x_3) \quad (30)$$

If this results is combined with the results of Equations 26 and 28, we can derive the following expression after

relabeling:

$$f(x_1, x_2, x_3) = f_3(x_3) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot f_2(x_2) \cdot c_{13|2}(F_{12}(x_1|x_2), F_{23}(x_2|x_3)) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot f_1(x_1) \quad (31)$$

Rearranging this expression gives:

$$f(x_1, x_2, x_3) = f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot c_{12}(F_1(x_1), F_2(x_2)) \cdot c_{23}(F_2(x_2), F_3(x_3)) \cdot c_{13|2}(F_{12}(x_1|x_2), F_{23}(x_2|x_3)) \quad (32)$$

This is the full decomposition of a three dimensional joint distribution into a cascade of pair copulas. Again, this representation is not unique. It can be proven that if  $d$  represents the number of dimensions, the number of different R-vines equals:

$$\binom{d}{2} \times (d-2)! \times 2^{\binom{d-2}{2}} \quad (33)$$

There are two difficulties that need to be overcome, when Using the expression 32 into practice. Firstly we know that due to relabeling there are many different ways to construct this equation. Therefore we need to specify a preferred structure to work with. This boils down to selecting a certain Vine structure which is addressed in Section 4.6.2. The other problem concerns the conditional data we obtain when decomposing a joint probability density function into pair copulas. In other words, the data has to be transformed to conditional data. To put the R-vine theory into practice these issues have to be tackled. Firstly we will address the problem of choosing an appropriate structures, secondly the issue of the conditional copulas and the conditional data will be addressed.

#### 4.6.2 Regular Vines (R-Vines)

Bedford & Cooke (2002) have introduced a graphical model denoted by *the regular vine* (R-vine) to help organizing different Vine structures. Regular vines are known from the graph theory. In this section the theoretical background of regular Vines is explained, while the graph theory is kept to a minimum.

*Graphs* are collections of nodes which are connected to each other by edges. These edges have weight. We represent our variables by nodes and their dependence in terms of Kendall's  $\tau$  by edges. *Graphs* can be expressed as follows:

**Definition:** (Graph). Let  $N$  be a certain set and let  $E$  be a subset of all possible two dimensional combinations of the elements in  $N$

$$E \subset \left\{ \{n_1, n_2\} \mid n_1, n_2 \in N \right\} \quad (34)$$

Then we the *nodes* are denoted by the elements of  $N$  and *edges* by the elements of  $E$ . A *graph*  $G$  is combination of edges and nodes:  $G = (N, E)$ .

**Definition:** (Path). A *path* is a trail in which all nodes are distinct except the first and last node. A trail is a walk in which all edges are distinct, which begins and ends with certain nodes. There is a start node and an end node, the nodes in between are called internal nodes.

**Definition:** (Cycle). A *cycle* is a path shares the same start and end node.

**Definition:** (Connected Graph). A *connected graph* is a graph  $G = (N, E)$  in which all possible nodes  $x, y \in N$  ( $x \neq y$ ) are connected by a path.

As mentioned before, R-vines split a multivariate distribution into a cascade of pair-copulas. These cascade consist of several different *trees*. In graph theory a tree can be defined as follows:

**Definition:** (Tree). A *tree* is a graph  $T = (N, E)$  that is connected and has no cycles.

Bedford & Cooke (2002) formulated R-Vines as being a subset of vines with the characteristics defined below.

**Definition:** A  $d$ -dimensional regular vine is a sequence of  $d - 1$  linked trees with nodes  $N_i$  and edges  $E_i$ .  $V = \{T_1, \dots, T_{d-1}\}$ .

1. Tree  $T_1$  has nodes  $N_1 = \{1, \dots, d\}$  and edges denoted by  $E_1$

2. For  $i \geq 2$  tree  $T_i$  has nodes  $N_i = E_{i-1}$  and edge set  $E_i$

Furthermore a vine  $V$  is considered to be a R-Vine on  $n$  elements if:

3. For  $i = 2, \dots, d - 1$  and  $\{a, b\} \in E_i$  with  $a = \{a_1, a_2\}$  and  $b = \{b_1, b_2\}$ , it must hold that  $|a \cap b| = 1$  (Proximity condition)

The proximity condition would be violated if there is an edge connecting node, while the connected nodes/edges do not share a common node in trees before.

There are two types of Vines that are used commonly. These are subsets of the R-vine and named: the canonical vine (C-vine) and D-vine. D-Vines and C-Vines are special cases of an R-Vine. A C-Vine is an R-Vine with each tree  $T_i, i = 1, \dots, n - 1$ , has one unique node of degree  $n - 1$ , while a D-Vine is an R-Vine if each node in  $T_1$  has degree of at most 2. For the C-vine the  $n$ -dimensional probability density function is stated as (Aas et al. 2009):

$$f(x_1, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{j,j+i|1, \dots, j-1}(F(x_j|x_1, \dots, x_{j-1}), F(x_{j+i}|x_1, \dots, x_{j-1})), \quad (35)$$

The graph structure of this particular type in the case five variables are used, is reported in the Appendix 9 along with a representation of the decomposition in Equation 50. It has to be kept in mind that these representations in the Appendix only refer to a particular C-vine. By using relabeling, other variations of the C-vine can be constructed. The C-vine basically links particular variable to all the remaining others. In other words, one variable will be placed in the center. Once all the links have been estimated by bivariate copulas, the dependences between the links their selves are estimated and so on. This process will be repeated until all possible conditional links are estimated.

The  $n$ -dimensional probability density function for the D-vine can be described as (Aas et al. 2009):

$$f(x_1, \dots, x_n) = \prod_{k=1}^n f(x_k) \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,i+j|i+1, \dots, i+j-1}(F(x_i|x_{i+1}, \dots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \dots, x_{i+j-1})) \quad (36)$$

The graph structure of this particular type for five variables is reported in the Appendix C in Figure 10, along with a representation of the decomposition in Equation 50. Again we have to hold in mind that there are many different other variations possible due to relabeling. Unlike the C-vine, the D-vine does not centralize particular variables. In the D-Vine variables are linked side by side. In other words, a non-looped chain of variables is created. Once these links have been estimated, the dependencies between the links their selves will be estimated, while keeping this chain structure.

Besides these two type of structures many more structures can be used. For instance Kurowicka (2010) proposed a method based on partial correlations to select an R-Vine. In Difmann (2010) a method is proposed based on Kendall's  $\tau$ . In this research we will use the same methodology as in Difmann (2010). Difmann (2010) assigns weights to the edges and chooses a sub graph <sup>13</sup>, which necessarily has to be a tree, in such a way that the sum of the weights of the edges are maximized. The weights between two nodes is defined as the empirical Kendall's  $\tau$

<sup>13</sup>A sub graph has the same nodes but not necessarily the same edges

between these particular variables. This particular sub graph is called the *Maximum Spanning Tree* or the *MST*. We use the Prim's Algorithm from Cormen et al. (2001) to determine the MST. Prim's Algorithm is originally described as a way to find minimum spanning tree, but it works in both ways if we choose to multiply the edge values by minus one.

Prim's Algorithm can be described as follows:

Step 1: Form the connected graph  $G = (N, E)$  and create an empty sub graph  $T = (N', E')$

Step 2: Start with an arbitrarily chosen node from the graph, say  $n^*$ .

Step 3:  $N' = \{n^*\}$

Step 4: **while**  $T$  does not contain all nodes in  $G$  **do**:

Step 5: Select an edge  $e \in E$  such that a new node is connected with  $E'$  and has the minimal edge value.

Step 6: Add this new node to  $N'$  and the new edge  $e$  to  $E'$

Step 7: **end while**

In Cormen et al. (2001) it is proven that the final tree  $T$  is a MST.

To select a Vine structure in our research we follow the methodology of Dißmann (2010). This means that we select our structure based on the tree that maximizes the sum of the edge values, where the edge values are represented by the empirical Kendall's  $\tau$ . Once we find the MST of the next tree, containing the MST edges of the previous tree as nodes. This process will be repeated until there is only one node left.

#### 4.6.3 Estimating Copulas

Where in Section 4.6.2 we show how to find an appropriate Vine structure, we show in this section how we choose a bivariate copula and how to estimate it. It is known that bivariate copulas can be estimated in several ways. One approach uses Maximum Likelihood estimation while others use Kendall's  $\tau$  or Spearman's  $\rho$  to estimate the copula parameter. In this research we choose to use Maximum Likelihood in order to estimate the copula parameters. We motivate our choice of estimation by two different reasons. Firstly Maximum Likelihood is more flexible when the number of copula variables are greater than one. Secondly Maximum Likelihood estimation is a standard in vine copula estimation Dißmann (2010) and Kurowicka (2010).

Let  $\Theta$  denote the set of copula parameters that need to be estimated. If for example a C-vine structure is assumed the n-dimensional log likelihood ( $l(\Theta)$ ) can be formulated as follows:

$$l(\Theta) = \sum_{j=1}^{n-1} \sum_{i=1}^{n-j} \sum_{t=1}^T \log \left( c_{j,j+i|1, \dots, j-1} (F(x_{j,t}|x_{1,t}, \dots, x_{j-1,t}), F(x_{j+i,t}|x_{1,t}, \dots, x_{j-1,t})) \right), \quad (37)$$

All marginal probability density function drop out of the likelihood function, since they are deterministic. This boils down maximizing the likelihoods of all the individual pair-copula (estimating  $\Theta$ ). Since it is infeasible to use Maximum Likelihood on the entire Vine structure, we estimate all the individual copula pairs in each tree separately. Therefore we assume that the copula parameters optima in the next tree are unaffected by the copula parameters in the previous trees.

In order to estimate the copula parameters in other trees than the first tree, we need to transform our data to conditional data. The data can be transformed to conditional data by using a h-function. We can define a h-function as follows:

$$h(x_i, x_j, \Theta_{ij}) = F(X_i|X_j)|_{X_i=x_i, X_j=x_j, \Theta_{ij}} = \left. \frac{\partial C_{ij}(F(X_i), F(X_j))}{\partial F(X_j)} \right|_{X_i=x_i, X_j=x_j, \Theta_{ij}}, \quad (38)$$

where  $h(x_i, x_j, \Theta_{ij})$  is the h-function always conditioned on the second parameter (in this case  $x_j$ ).  $\Theta_{ij}$  stands for the bivariate copula parameter between variable  $i$  and variable  $j$ . It is important to hold in mind that due to the fact that some bivariate copulas are not symmetric (this means that  $C_{12}(u1, u2) \neq C_{12}(u2, u1)$ ). Therefore, the same partial derivative function that is being used for conditioning  $j$  on  $i$  may not generally be used for the case that  $i$  is conditioned on  $j$ . In other words the h-function is unique for the symmetric bivariate copula and differs for asymmetric bivariate copulas.

The beauty of using graph theory in a copula context is that the results of the previous tree are used to obtain the conditional data in the next tree using h-functions. In order to illustrate this, we use the C-Vine from expression 35. We now consider the fourth and the fifth line of this expression. Firstly  $c_{34|12}(F(x_3|x_1, x_2), F(x_4|x_1, x_2))$  is considered. To estimate this copula, expressions of  $F(x_3|x_1, x_2)$  and  $F(x_4|x_1, x_2)$  are required in order to obtain the conditional data. In equation 39 these two conditional cumulative density functions are described using the appropriate h-functions.

$$\begin{aligned} F(x_3|x_1, x_2) &= h(F(x_3|x_2), F(x_1|x_2), \Theta_{31}) = h(h(x_3, x_2, \Theta_{32}), h(x_1, x_2, \Theta_{12}), \Theta_{31}) \\ F(x_4|x_1, x_2) &= h(F(x_4|x_1), F(x_2|x_1), \Theta_{42}) = h(h(x_4, x_1, \Theta_{41}), h(x_2, x_1, \Theta_{21}), \Theta_{42}) \end{aligned} \quad (39)$$

#### 4.6.4 Sampling from Vine Copulas

In the end we want to sample from the estimated Vine in order to obtain VaR or ES estimates. In order to sample from the Vine the inverse h-function is required ( $h^{-1}$ ). The  $h^{-1}(x_i, x_j, \Theta_{ij})$  is the inverted h-function with respect to the first variable (in this case  $x_i$ ) or equivalently the inverse of the conditional distribution function. This can be expressed as:

$$F(x_i|x_j) = h(x_i, x_j, \Theta_{ij}) = c \quad \Leftrightarrow \quad F^{-1}(c|x_j) = h^{-1}(c, x_j, \Theta_{ij}) = x_i \quad (40)$$

where  $c$  is a certain constant. Simulation from a Vine is briefly discussed in Bedford & Cooke (2002) and Kurowicka & Cooke (2006). Aas et al. (2009) gives a practical approach on how to sample from C-Vines and D-Vines. The basic idea of sampling from a Vine is shown by an example. In this example sampling from an d-dimensional C-Vine is considered.

Step 1: Sample  $n$  independent uniform  $[0,1]$  variables,  $w_1, \dots, w_n$

Step 2: Set  $x_1 = w_1$

Step 3: Assume that the copulas are symmetric. Now, use that  $C_{12}$  has already been estimated in tree two and therefore the corresponding h-function is known, to obtain  $x_2 = F_{2|1}^{-1}(w_2|x_1) = h^{-1}(w_2, x_1, \Theta_{12})$

Step 4: Proceed with the next tree and use the estimated copula  $C_{23|1}$  to obtain a sampled value of  $x_3$  as follows:

$$\begin{aligned} F(x_3|x_1, x_2) &= h(F(x_3|x_1), F(x_2|x_1), \Theta_{32}) = h(h(x_3, x_1, \Theta_{31}), h(x_2, x_1, \Theta_{21}), \Theta_{32}) = w_3 \Leftrightarrow \\ h^{-1}(w_3, h(x_2, x_1, \Theta_{21}), \Theta_{32}) &= h(x_3, x_1, \Theta_{31}) \Leftrightarrow x_3 = h^{-1}(h^{-1}(w_3, h(x_2, x_1, \Theta_{21}), \Theta_{32}), x_1, \Theta_{31}) \end{aligned}$$

Step 5: Proceed until all  $n$  variables are sampled one time.

Using the same reasoning it is possible to sample from all other possible Vines.

#### 4.6.5 Sequential Method

In our research we follow the Vine selection procedure from Difsmann (2010). He refers to this procedure as the *sequential method*. This method uses the Kendall's  $\tau$  for selecting a Vine structure. This method is motivated by the idea that the strongest dependencies need to be captured in the first trees, since these are the most important ones. Furthermore this approach will avoid numerical errors in higher order trees (Difsmann (2010), Schepsmeier & Czado (2015) and Aas et al. (2009)). Higher order trees contain many very weak relationships, which are less

inclined to estimation errors, since copula are generally speaking very capable of modeling (close to) independence.

The sequential algorithm can be described in several different steps.

Step 1: Construct a tree  $T_1$  with  $d$  nodes and with all possible edges.

Step 2: Compute the absolute empirical Kendall's  $\tau$  for all the possible pairs (number of all possible pairs equals the number all possible edges).

Step 3: Use Prim's Algorithm (Prim 1957) to determine the maximum spanning tree.

Step 4: For all the pairs found by Prim's algorithm select a certain copula and fit this copula to the data by using Maximum Likelihood Estimation.

Step 5: Proceed by constructing a tree, where the nodes are represented by the edges in the previous tree.

Step 6: Transform the observations to conditional data (using the  $h$ -function).

Step 7: Compute the absolute empirical Kendall's  $\tau$  for all the possible pairs of conditional data.

Step 8: Proceed with step 3 until the R-Vine is fully specified. If there is at a certain tree only one option to create a spanning tree, proceed by using the D-Vine structure.

An example of the sequential method for a joint distribution of six different assets is given in the Appendix C in Figure 20.

#### 4.6.6 Model Comparison of R-Vines

The comparison of R-Vines is useful to evaluate the performance of different vine structures. We can obtain different vine structures when we use:

- Different tree structures in Vines
- Different copula families for the pairs in the edges
- Both different tree structures in vines and different copula families for the pairs

Because different R-Vines are considered to be non-nested, the test of Vuong (1989) for non-nested models can be applied based on the KullBack-Leibler information criteria (Clarke 2007). Consider two different models that try to model the true model  $h_0$ ,  $F_\beta = f(Y|\beta)$  and  $G_\gamma = g(Y|\gamma)$ . Let  $\beta_*$  and  $\gamma_*$  be the true parameters. The null hypothesis of the test is that the two different models are equally close to the true specification:

$$H_0 : E_0 \left[ \log \frac{f(Y|\beta_*)}{g(Y|\gamma_*)} \right] = 0 \quad (41)$$

Vuong (1989) proves that under general conditions  $(1/n)$  ( $n$  stands for the number of observations) times the likelihood ratio statistic approximates this expected value:

$$\frac{1}{n} LR(\hat{\beta}, \hat{\gamma}) \rightarrow E_0 \left[ \log \frac{f(Y|\beta_*)}{g(Y|\gamma_*)} \right], \quad \text{if } n \rightarrow \infty \quad (42)$$

Let  $\hat{\beta}$  and  $\hat{\gamma}$  are the maximum likelihood estimator of  $\beta_*$  and  $\gamma_*$ .  $LR(\hat{\beta}, \hat{\gamma}) = \hat{L}_f - \hat{L}_g$ , where  $\hat{L}_f$  and  $\hat{L}_g$  are the maximized likelihoods. The resulting likelihood ratio statistic is asymptotically normally distributed and the test statistic is under the null hypothesis:

$$\hat{V}_n := \frac{LR(\hat{\beta}, \hat{\gamma})}{(\sqrt{n})\hat{\omega}} \rightarrow N(0, 1) \quad (43)$$

with  $\hat{\omega}$  is the estimated standard deviation:

$$\hat{\omega} = \frac{1}{n} \sum_{i=1}^n \left[ \log \frac{f(Y|\beta_*)}{g(Y|\gamma_*)} \right]^2 - \left[ \frac{1}{n} \sum_{i=1}^n \log \frac{f(Y|\beta_*)}{g(Y|\gamma_*)} \right]^2 \quad (44)$$

The corresponding p-value is  $2(1 - \Phi(|\hat{V}_n|))$ , where  $\Phi$  is the normal cumulative density function. Not rejecting the  $H_0$  means that both models are considered to be significantly equally good. However if the  $H_0$  does not hold then model  $F_\beta$  is in favor when the value of  $\hat{V}_n$  is positive. If this value is negative then the model  $G_\gamma$  is in favor.

## 4.7 Multivariate Extreme Value Theory

Besides copulas we use another way to model the joint distribution. In this we show how multivariate extreme value theory (MEVT) is used to model the dependence between variables. This Section will be split up into three different parts. First we focus on the basics of this MEVT model. Next we show the estimation procedure of this model and finally we elaborate on how to sample from this model.

### 4.7.1 MEVT Model

The basic idea of MEVT is to model the joint tail behavior of multiple assets. Since VaR and ES on the extreme upper or lower quantiles, a model solely based on these certain parts could improve VaR and ES estimations. The MEVT model we use is related with the exceedances over threshold models. The general distribution function of univariate exceedance over threshold models can be described as:  $F_{\text{excess}} = \Pr[X - u \leq x | |X| > u]$ , where  $u$  is a single threshold for both upper and lower bounds,  $X$  a random variable and  $x$  a constant. To extent this theory to more dimensions, say  $d$  dimensions, many different extreme value distributions can be constructed. We can for example construct  $F_{\text{excess}} = \Pr[X_i - u \leq x | |X_j| > u]$ ,  $F_{\text{excess}} = \Pr[X_i - u \leq x | |X_k| > u, |X_j| < u]$ , or  $F_{\text{excess}} = \Pr[X_i - u, X_j - u \leq x | |X_k| > u]$  and so on. The number of scenarios increases even more if for example time lags are added or when the order of variables is changed. Zhang & Huang (2006) tried to tackle this issue by constructing a three-state Markov chain to account for some of these scenarios. In other words, including lots of dimensions in the model leads to great problems.

Hilal et al. (2014) divides the problem in less scenarios and make POT multivariate EVT feasible from a practical point of view. Hilal et al. (2014) only accounts for the conditional probability of one variable given another certain variable is extreme (positive or negative). These models  $\mathbf{M}$  are denoted by  $\mathbf{M}_{j|i}^\dagger, \forall j, i$  which  $j \neq i$ , where  $\dagger \in (+, -)$  represents if the observation conditioned on is extremely positive or negative. The number different dependence models now equals  $d \times (d - 1) \times 2$ . McNeil et al. (2005) show that autocorrelation of extreme returns is not present or very short-lived events. To keep the model as simple as possible and because we do not wish to over parameterize the model, these effects will not be captured by the model.

The model from Hilal et al. (2014) requires i.i.d. data as input. The ARMA-MS-GARCH-GJR filter will be used in order to obtain data that is (close to) i.i.d.. Moreover, this methodology requires the marginals to be Laplace transformed (in order to obtain the correct standardizing functions). Let  $\hat{F}_{X_i}(X_{i,t})$  be the  $t^{\text{th}}$  observation of variable  $i$  fitted accordingly to the appropriate marginal estimated cumulative density function of variable  $i$ , then

$$Y_{i,t} = F_Y^{-1}[\hat{F}_{X_{i,t}}(X_{i,t})] = \begin{cases} \log[2\hat{F}_{X_{i,t}}(X_{i,t})] & \hat{F}_{X_{i,t}}(X_{i,t}) \leq 0.5, \\ -\log[2(1 - \hat{F}_{X_{i,t}}(X_{i,t}))] & \hat{F}_{X_{i,t}}(X_{i,t}) > 0.5 \end{cases} \quad (45)$$

is the Laplace transformation of a uniform random variable. The Laplace distribution is known for its exponential upper and lower tails.

The main objective of Hilal et al. (2011) and Hilal et al. (2014) is to capture the dependence structure between  $Y_{i,t}$  and  $Y_{j,t}$  for both the upper and lower tail. We use the modeling assumption from Heffernan & Tawn (2004) that there exist some standardizing constants  $a_{j|i}^\dagger \in [-1, 1]$  and  $b_{j|i}^\dagger \in (-\infty, 1)$  such that for all fixed  $z_{j|i,t}^\dagger = z$  the following condition holds:

$$\lim_{y_i \rightarrow \infty} \left\{ \Pr[Y_{j|i} \leq a_{j|i}^\dagger(Y_i) + b_{j|i}^\dagger(Y_i)z \mid Y_i = y_i] \right\} = G_{j|i}(z)^\dagger, \quad (46)$$

where  $G_{j|i}(z)^\dagger$  is a non-degenerate function and  $\lim_{z \rightarrow \infty} G_{j|i}(z)^\dagger = 1$ . The standardizing constants can take on numerous of forms (a detailed table about different standardizing constants can be found in Heffernan & Tawn (2004)). For the Laplace margins the appropriate standardization functions are:

$$\begin{aligned} a_{j|i}^{(l)\dagger}(Y_i) &= a_{j|i}^\dagger \dagger Y_i \\ b_{j|i}^{(l)\dagger}(Y_i) &= \dagger Y_i^{b_{j|i}^\dagger} \end{aligned}$$

Using this model the standardized residuals can be obtained:

$$Z_{j|i,t}^{(l)\dagger} = \frac{Y_{j,t+l} - a_{j|i}^\dagger \dagger Y_{i,t}}{\dagger Y_{i,t}^{b_{j|i}^\dagger}}$$

Using this standardized variable, equation 46 can be rewritten as:

$$\lim_{y_i \rightarrow \infty} \left\{ \Pr[Z_{j|i,t}^\dagger \leq z \mid Y_{i,t} = y_i] \right\} = G_{j|i}(z)^\dagger, \quad (47)$$

The following relationship holds given that  $\dagger Y_i \rightarrow \infty$ :

$$Y_{j,t+l} = a_{j|i}^\dagger \dagger Y_{i,t} + \dagger Y_{i,t}^{b_{j|i}^\dagger} Z_{j|i,t}^\dagger$$

Using this relationship three different categories of asymptotic dependence can be specified.

1. Positive asymptotic dependence if  $0 < a_{j|i}^\dagger < 1$  or **both**  $a_{j|i}^\dagger = 0$  **and**  $b_{j|i}^\dagger > 0$ .
2. Negative asymptotic dependence if  $-1 < a_{j|i}^\dagger < 0$ .
3. Close to asymptotic independence if **both**  $a_{j|i}^\dagger = 0$  **and**  $b_{j|i}^\dagger \leq 0$

In the remainder we discuss on how to estimate the model and how to sample from it.

#### 4.7.2 MEVT estimation

In order to estimate this model, the estimation method from Heffernan & Tawn (2004), Hilal et al. (2011) and Hilal et al. (2014) is used. All the thresholds  $u_{i|j}^\dagger$  are assumed to be the same and correspond to the 5<sup>th</sup> percentile of the Laplace distribution for the lower threshold ( $u_{i|j}^- = u^-$ ) and the 95<sup>th</sup> percentile for the upper threshold ( $u_{i|j}^+ = u^+$ ). Let  $u_{Y+}$  denote the general upper threshold and  $u_{Y-}$  the general lower threshold. To estimate the model a distributional assumption has to be made. The standardizing functions are assumed to be normalizing in the sense that  $Z_{j|i}^\dagger \sim N(\mu_{j|i}^\dagger, \sigma_{j|i}^{2\dagger})$  when  $\dagger Y_i \rightarrow \infty$ .  $\mu_{j|i}^\dagger$  is the finite mean and  $\sigma_{j|i}^\dagger$  the standard deviation. Using this distributional assumption the models  $\mathbf{M}_{j|i}^\dagger$  can be estimated using maximum likelihood estimation. The log likelihood ( $\log L$ ) function that needs to be maximized with respect to  $\Theta_{j|i}^\dagger$  is stated as follows:

$$\log L(\Theta_{j|i}^\dagger) = - \sum_{t=1}^{n_{j|i}^\dagger} \left[ \log \left\{ \sigma_{j|i}^\dagger \dagger Y_{i(t)}^{b_{j|i}^\dagger} \right\} + \frac{1}{2} \left\{ \frac{Y_{j(t)} - a_{j|i}^\dagger \dagger Y_{i(t)} - \mu_{j|i}^\dagger \dagger Y_{i(t)}^{b_{j|i}^\dagger}}{\sigma_{j|i}^\dagger \dagger Y_{i(t)}^{b_{j|i}^\dagger}} \right\}^2 \right],$$

where  $\Theta_{j|i}^\dagger = [\mu_{j|i}^\dagger, \sigma_{j|i}^\dagger, a_{j|i}^\dagger, b_{j|i}^\dagger]$

#### 4.7.3 MEVT Sampling

Once the models are fitted, we can start sampling observations from the models. The target is to sample a  $d$  dimensional vector of returns. In order to do so, we consider  $2d + 1$  different regions or intervals:

$$\begin{aligned}
\mathbb{R}_1^+ &= \left\{ Y_1 > u_{Y+} \quad \cap \quad Y_1 \geq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{1\} \right\} \\
\mathbb{R}_2^+ &= \left\{ Y_2 > u_{Y+} \quad \cap \quad Y_2 \geq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{2\} \right\} \\
&\vdots \\
\mathbb{R}_d^+ &= \left\{ Y_d > u_{Y+} \quad \cap \quad Y_d \geq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{d\} \right\} \\
\mathbb{R}_1^- &= \left\{ Y_1 > u_{Y-} \quad \cap \quad Y_1 \leq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{1\} \right\} \\
\mathbb{R}_2^- &= \left\{ Y_2 > u_{Y-} \quad \cap \quad Y_2 \leq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{2\} \right\} \\
&\vdots \\
\mathbb{R}_d^- &= \left\{ Y_d > u_{Y-} \quad \cap \quad Y_d \leq Y_I, \quad \forall I \in \{1, \dots, d\} \setminus \{d\} \right\} \\
\mathbb{R}_0 &= \left\{ Y_I \leq u_{Y+} \quad \cap \quad Y_I \geq u_{Y-}, \quad \forall I \in \{1, \dots, d\} \right\}
\end{aligned}$$

This is a natural way to split up the space (Hilal et al. 2011). The associated possibilities of observations of vector  $Y$  falling in the particular region are:

$$\begin{aligned}
P_1^+ &= \Pr(Y \in \mathbb{R}_1^+) = \Pr\left[Y \in \mathbb{R}_1^+ \mid Y_1 > u_{Y+}\right] \Pr(Y_1 > u_{Y+}) \\
P_2^+ &= \Pr(Y \in \mathbb{R}_2^+) = \Pr\left[Y \in \mathbb{R}_2^+ \mid Y_2 > u_{Y+}\right] \Pr(Y_2 > u_{Y+}) \\
&\vdots \\
P_d^+ &= \Pr(Y \in \mathbb{R}_d^+) = \Pr\left[Y \in \mathbb{R}_d^+ \mid Y_d > u_{Y+}\right] \Pr(Y_d > u_{Y+}) \\
P_1^- &= \Pr(Y \in \mathbb{R}_1^-) = \Pr\left[Y \in \mathbb{R}_1^- \mid Y_1 > u_{Y-}\right] \Pr(Y_1 > u_{Y-}) \\
P_2^- &= \Pr(Y \in \mathbb{R}_2^-) = \Pr\left[Y \in \mathbb{R}_2^- \mid Y_2 > u_{Y-}\right] \Pr(Y_2 > u_{Y-}) \\
&\vdots \\
P_d^- &= \Pr(Y \in \mathbb{R}_d^-) = \Pr\left[Y \in \mathbb{R}_d^- \mid Y_d > u_{Y-}\right] \Pr(Y_d > u_{Y-}) \\
P_0 &= \Pr(Y \in \mathbb{R}_0) = 1 - \sum_{i=1}^d (P_i^- + P_i^+)
\end{aligned}$$

The probability can be approximated using Monte Carlo simulation. The algorithm for a certain  $P_i^+$  is described by the following steps:

### Algorithm 1a

Step 1: Simulate an extreme positive  $\hat{Y}_i^+ = u_Y^+ + E$ , where  $E \sim \text{Exponential}(1)$  using the fact that both tails of the Laplace transformed variable are exponential.

Step 2: Sample a  $d - 1$  vector  $\hat{Z}_{j|i}^+$  independently from  $\hat{Y}_i$  using the estimated models  $\mathbf{M}_{j|i}^+$ .

Step 3: Obtain the simulated  $\hat{Y}_j^+ = \hat{a}_{j|i}^+ \cdot \hat{Y}_i + \hat{Y}_i^{\hat{b}_{j|i}^+} \cdot \hat{Z}_{j|i}^+$

Step 4: Repeat step 2 and 3 for all the possible  $j \in \{1, \dots, d\} \setminus \{i\}$

Step 5: Let  $Y_{|i}$  denote  $(Y_{1|i}, Y_{2|i}, \dots, Y_{d|i})$ . Now the probability can be estimated using Monte Carlo:

$$P_i^+ \approx \left\{ \frac{1}{K} \sum_{k=1}^K \mathbb{I}[(\hat{Y}_{|i}^+)^{(k)} \in \mathbb{R}_I^+] \right\} \left\{ \frac{1}{2} \exp(-u_{Y^+}) \right\}$$

The probability can be approximated using Monte Carlo. The algorithm for a certain  $P_i^-$  is described by the following steps:

### Algorithm 1b

Step 1: Simulate an extreme positive  $\hat{Y}_i^+ = u_Y^- - E$ , where  $E \sim \text{Exponential}(1)$  using the fact that both tails of the Laplace transformed variable are exponential.

Step 2: Sample a  $d - 1$  vector  $\hat{Z}_{j|i}^-$  independently from  $\hat{Y}_i$  using the estimated models  $\mathbf{M}_{j|i}^-$ .

Step 3: Obtain the simulated  $\hat{Y}_j^- = \hat{a}_{j|i}^+ \cdot \hat{Y}_i + \hat{Y}_i^{\hat{b}_{j|i}} \cdot \hat{Z}_{j|i}^-$

Step 4: Repeat step 2 and 3 for all the possible  $j \in \{1, \dots, d\} \setminus \{i\}$

Step 5: Let  $Y_{|i}$  denote  $(Y_{1|i}, Y_{2|i}, \dots, Y_{d|i})$ . Now the probability can be estimated using Monte Carlo:

$$P_i^- \approx \left\{ \frac{1}{K} \sum_{k=1}^K \mathbb{I}[(\hat{Y}_{|i}^-)^{(k)} \in \mathbb{R}_I^+] \right\} \left\{ \frac{1}{2} \exp(-u_{Y^-}) \right\}$$

Combining the results from algorithm 1a and 1b and using that the sum of all the probability has to be equal to one, we can obtain the following expression for  $P_0$ :

$$P_0 = 1 - \left( \sum_{i=1}^d P_i^+ + P_i^- \right)$$

Once the probabilities are estimated, sampling from the model is made possible. The sampling algorithm can be described as follows:

### Algorithm 2

Step 1: Sample  $[k_0, k_1^+, \dots, k_d^+, k_1^-, \dots, k_d^-]$  from a multinomial distribution with estimated probabilities

$\hat{P}_0, \hat{P}_1^+, \hat{P}_2^+, \dots, \hat{P}_d^+, \hat{P}_1^-, \hat{P}_2^-, \dots, \hat{P}_d^-$ . Where  $k_i^\dagger$  represents the number of cases that the vector that needs to be sampled lies in region  $\mathbb{R}_i^\dagger$  for  $\dagger \in \{+, -\}$ .

Step 2: Sample  $k_0$  empirically from the region  $\mathbb{R}_0$

Step 3: Sample  $\hat{Y}_{|1}^+$  using the algorithm 1a and 1b until  $k_1^+$  are accepted in the region  $\mathbb{R}_1^+$

Step 4: Repeat Step 3 for all different  $i = 1, \dots, d$  and  $\dagger = \{+, -\}$

## 4.8 Backtesting

Historical data is used in order to evaluate the performance of the different models when estimating the VaR or the ES. The accuracy of the VaR estimation can be tested using the conditional or unconditional coverage tests. These tests count the number of times the VaR is exceeded, i.e. when portfolio losses exceed VaR estimates. If the number of exceedances is significantly greater than the selected confidence level corresponding with the VaR then the estimated VaR is considered to underestimate the risks. On contrary, when the number of exceedances is significantly smaller than the selected confidence level, then the VaR is considered to overestimate the risks. The exceedances often occur in clusters (for example during financial distress). In theory the exceedances in a good VaR model occur independently over time. This independence feature is not taken into account when conducting an unconditional coverage test, unlike the conditional coverage test (Christoffersen 1998). The testing can be conducted in two different ways. Firstly there static models can be used. In these static models the VaR is kept

fixed over time. The advantage of this method is its simplicity, there is no requirement to estimate the VaR repeatedly in a window. The other method is the dynamic method. Then a window is used to estimate all the different models over and over again. In our research we prefer the dynamic models.

We can also backtest the ES. We elaborate on this test in Section 4.8.2

#### 4.8.1 POF-test

The POF-test from Kupiec (1995) is an unconditional coverage test. To conduct this test  $N$  days in the future are collected. Now, let  $\pi$  denote the probability that observation  $i$  in the future exceeds the  $\text{VaR}_a$ . Assuming that this probability is equal the likelihood ratio is given by:

$$L = \prod_{i=1}^N \pi^{I_i} (1 - \pi)^{1-I_i} = \pi^n (1 - \pi)^{N-n},$$

where  $I_i$  is an indicator function that is one if the  $\text{VaR}_a$  is exceeded and zero otherwise and  $n = \sum_{i=1}^N I_i$ . Using maximum likelihood estimation would give:  $\hat{\pi} = N/n$ . Let the null hypothesis be that the  $\text{VaR}_a$  describes the data well and the alternative hypothesis be that the  $\text{VaR}_a$  fails. Under the null hypothesis we get the likelihood:

$$L_0 = (1 - \pi)^n \pi^{N-n},$$

while under the alternative hypothesis we have:

$$L_1 = \left(\frac{n}{N}\right)^n \left(1 - \frac{n}{N}\right)^{N-n},$$

Now the log likelihood ratio can be formulated as:  $-(\log L_0 - \log L_1)$  which is  $\chi^2$  distributed with one degree of freedom.

#### 4.8.2 ES Zero Mean Test

McNeil & Frey (2000) propose a way to backtest the Expected Shortfall (ES). The test that is being used is based on the idea that when a certain return exceeds the VaR, given that the VaR is specified correctly, then the exceeded loss or profit minus the ES should have mean zero. This can be described as follows:

$$S_t = \frac{(R_t - ES_\alpha)}{\sigma_t} I_{\{R_t > \text{VaR}_\alpha\}}, \quad \text{then}$$

$$H_0 : \frac{1}{T} \sum_{t=1}^T S_t = 0$$

To test if this null hypothesis holds we can either use the standard one sample t test, with the assumption that  $S_t$  is i.i.d. distributed, or use the bootstrap test from Efron & Tibshirani (1994), which does not require a distributional assumption. In this case we use the one-sided t test with null hypothesis that the excess shortfall has zero mean against the alternative hypothesis that the excess shortfall has mean greater than zero, so that the shortfall is systematically underestimated.

## 5 Results

In this section we provide the empirical results of applying the described methods to a dataset containing seven different major world indices. We decide to keep the same structure as in the methodology. Therefore we split the result reporting roughly into two different parts. In the first part we elaborate on the results of the marginal modeling and the two-step method (ARMA-MS-GARCH-GJR and semi-parametric modeling). In the second part

we analyze the dependence modeling using Vine copula and a joint distribution based on MEVT. We backtest these models in both a period of financial distress and a stable period.

## 5.1 Marginal Results

The marginal modeling consists of the two-step method containing the ARMA-MS-GARCH-GJR as first step and the semi-parametric modeling as second step.

### 5.1.1 ARMA-MS-GARCH-GJR

The estimated ARMA-MS-GARCH-GJR will operate as a filter for the return series. The target is that some of the stylized facts of returns reported by (McNeil et al. 2005) are filtered out. The ARMA part (if necessary) will focus on the autocorrelation of returns, the GARCH part the volatility clustering, the GJR part to capture the asymmetric behavior of the volatility process. Finally, we govern our returns series by an unobserved Markov process containing two different states. Firstly the number of lags and the number of regimes that are included need to be calibrated. This corresponds to calibrating  $i, j, k$  and  $l$  of the ARMA(i,j)-GARCH(k,l)-GJR(m), where the number of lags of the GJR is assumed to be equal to the number of ARCH lags ( $m = l$ ).

The number of regimes is set to two for different reasons. Firstly, the Financial Crisis in 2008 is part of our dataset, we want to model a volatility process that mainly distinguishes between crisis and stability. Secondly, we want to avoid overfitting. Adding states will add  $i + j + k + 2 \times l$  ARMA-GARCH-GJR parameters and will extent the transition matrix with transition probabilities. Thirdly, since for the backtesting we use a moving window setting in which we choose to re-estimate the parameters instead of keeping them fixed, adding many different states pushes the computation time up severely.

We choose the number of parameters  $i, j, k$  and  $l$  by comparing BIC values following Brockwell & Davis (1991). This lag selection occurs only at one period in time to avoid long computation time. It turns out that using this criterion for all series no ARMA effects and one GARCH lag is used. This results to a two state ARMA(i,j)-MS-GARCH(k,l)-GJR(m) with:  $i = 0, j = 0, k = 1$  and  $l = m = 1$  (Appendix B) Table 10. As expected given the stylized facts from McNeil et al. (2005) no ARMA lags are included. This means that the addition does not significantly improve the model fit in terms of BIC values.

The performance of the filter can be analyzed using autocorrelation plots. In the Appendix C in Figures 11-17 several autocorrelograms are shown for the filtered (standardized) and non-filtered series for all of the different marginals of the indices dataset. We observe a strong presence of autocorrelation in non-filtered absolute and squared returns. On the other hand the autocorrelation in the returns series is as expected rather weak. The standardized or filtered series hardly exhibit any autocorrelation in all of the normal, the absolute and squared cases. Using the Ljung-Box test and the Augmented Dickey-Fuller test we can test for i.i.d. (stationarity) of the filtered series. The results of these tests are reported in Appendix B in Table 7. This table gives us an indication that the filtered marginal returns seem to be very close to i.i.d.<sup>14</sup> The fitted student t distributions along with the normal distributions of all the marginals' residuals are shown in the Appendix C in Figure 18. It shows that the fitted student t distributions are very close to the normal distribution, which is no surprise given the high number of degrees of freedom (Table 5). For illustrating purposes the ARMA-MS-GARCH-GJR estimates over a period starting in 2002-07-08 until 2009-06-29 (containing 1500 observations) are shown in Table 5 below.

To illustrate the performance of the Markov switching of the GARCH-GJR model we show the returns over a certain period in time of the SMI along with the filtered probability of being in state 2 on a scale zero to one in gray in Figure 3. Furthermore we also choose to add the Viterbi path in this figure. In the Appendix C in Figure 19 these figures are shown for all the other indices as well. We choose to add the Viterbi path for illustration purposes. In the end we use the filtered path instead. The Viterbi path is a nice way to make the unobserved states observable by assigning the most likely sequence of unobserved states. The filtered path

---

<sup>14</sup>It has to be taken in mind that this is only an indication at one point in time. As we work with a moving window and recalibration of the models this does not have to be necessarily the case at other points in time. The number of lags is kept fixed over time due to limiting computation power.

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
$\omega_1$	0,00	0,00	0,00	0,00	0,00	0,00	0,00
$\alpha_1$	<i>0,00</i>	<i>0,00</i>	<i>0,00</i>	0,03	0,00	<i>0,00</i>	<i>0,00</i>
$\gamma_1$	0,20	0,14	0,12	0,19	0,57	0,16	0,16
$\beta_1$	0,78	0,83	0,86	0,85	0,72	0,83	0,90
$\omega_2$	0,00	0,00	0,00	0,00	0,00	0,00	0,00
$\alpha_2$	<i>0,00</i>	<i>0,00</i>	<i>0,00</i>	0,00	0,03	0,02	<i>0,00</i>
$\gamma_2$	0,12	0,10	0,11	0,06	0,06	0,10	0,18
$\beta_2$	0,92	0,93	0,92	0,94	0,91	0,91	0,83
P11	1,3%	0,8%	0,2%	0,1%	88,7%	0,3%	0,1%
P21	99,8%	99,8%	99,9%	99,9%	95,6%	99,8%	99,8%
DOF	27,73	29,49	29,50	9,38	97,79	52,92	17,18

Table 5: This table contains the estimated ARMA-MS-GARCH-GJR parameters using a window of 1500 daily observations starting in 2002-07-08 until 2009-06-29. The italic values are not significantly different from zero using a significance level of 1%.

however, gives probabilities of being in certain states. If the filtered probability is multiplied by the transition matrix the estimated probability of being in a certain state can be evaluated. The Viterbi path selects the highest estimated probability.

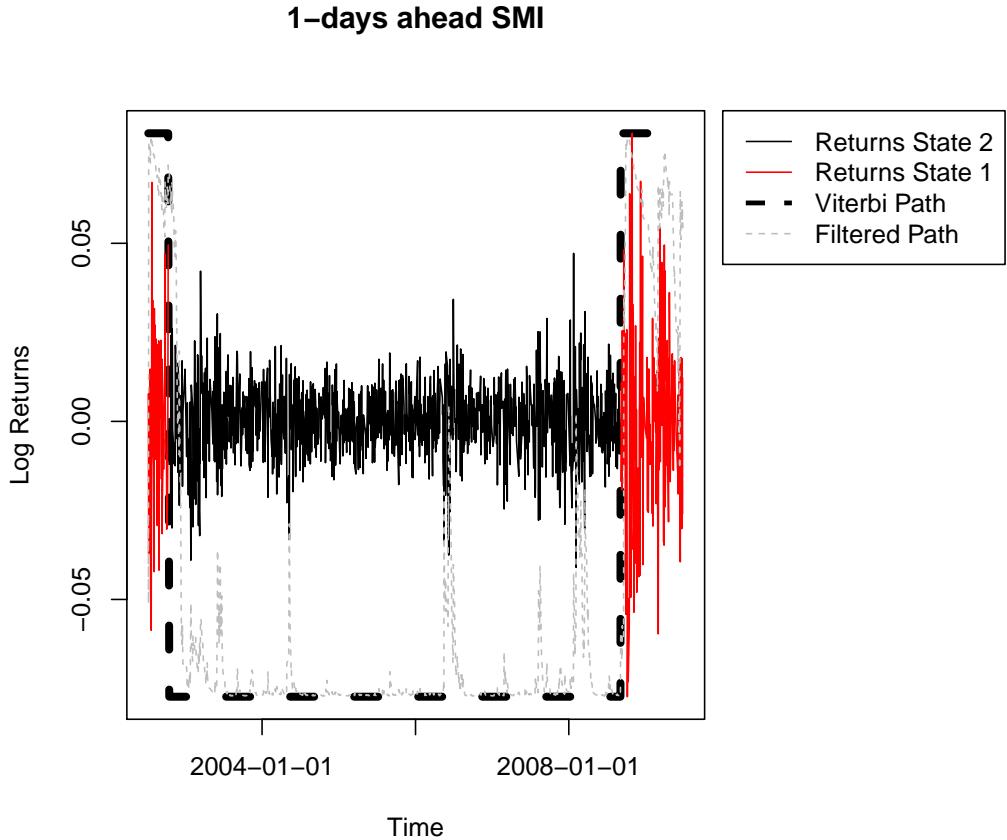


Figure 3: This graph contains the daily log returns for the SMI starting in 2002-07-08 until 2009-06-29, containing 1500 observations. The black and the red lines correspond with the actual returns. The dashed line represents the Viterbi path. When the Viterbi path is high then the Viterbi path is in state 2 and when it is low in state 1. The gray line represents the filtered path probability scaled to the maximum and the minimum log return.

We observe in figure 3 that there is a very high probability of being in state 1 until approximately end 2002 and after approximately mid 2008. In between there is a big probability of being in state 2. There is a clear indication that there is a very big probability of being in state 1 during periods of financial distress. We see that there is a very high probability of being in state 1 during the dotcom bubble burst in 2002. Furthermore we observe that this probability is also high during Financial Crisis, which hit its peak in mid 2008. If we combine this figure with table 5, we see that the estimated parameters of the GARCH-GJR model for the SMI differ across different states. We observe a higher value for  $\gamma$  in state 1 and a higher value for  $\beta$  in state 2. This indicates that there is an increased leverage effect in state 1 and an increased persistence in the volatility in state 2 given the volatility process:  $\sigma_t^2 = \omega + (\alpha_1 + \gamma_1 \cdot I_{t-1}) \cdot \epsilon_{t-1}^2 + \beta_1 \cdot \sigma_{t-1}^2$ . In other words the GARCH-GJR model is more reactive to shocks in state 1 than in state 2. This is in line with the statement from McNeil et al. (2005) that extreme returns (shocks) appear in clusters. Negative shocks result in high values of  $\epsilon^2$  and will have a larger impact on the estimated volatility in state 1 than in state 2 given the value of  $\gamma$ . The same conclusion can be drawn from the other indices. Only the NIKKEI225 seems to behave differently. This can be explained by the Lost Decade<sup>15</sup> of Japan. The Lost Decade is recognized as a period of economic stagnation due to the asset price bubble's

<sup>15</sup>The term originally refers to 1991 to 2000, but recently the decade from 2001 to 2010 is also included.

collapse in the early 90's.

### 5.1.2 Marginal Extreme Value Theory

In the Appendix C in Figures 20 and 21 the results of the threshold selection algorithm, or the automated eye-ball test in Section 4.4, are shown for indices dataset for the lower and the upper tails respectively for 1500 observations from 18-07-2000 until 01-07-2007. In order to avoid including too many observations in the tails, the  $w$  is set to 5 and the  $\epsilon$  is set to 0.1 and the  $h$  is set to 80%. The graphs show that the algorithm is quite capable of locating the first stable region, while keeping the number of observations in the tails quite low.

## 5.2 Multivariate Results

All of the marginal models will be modeled jointly by using Vine copula, MEVT or t-copula. The t-copula is used as benchmark model, since it is widely used in finance. In order to test the appropriateness of our methods regarding VaR and ES estimation, backtesting is used on both a equal weighted portfolio and a minimum variance portfolio. For the backtesting two different time periods will be used. Firstly we will examine the performance of our models during a period of financial distress (2007-07-01 until 2009-07-01 containing 389 observations) and secondly during a more stable time period (2015-01-01 until 2017-01-01 containing 391 observations). A rolling window will be used containing 1500 observations. In Section 5.3 the estimation of a certain Vine copula will be discussed, in Section 5.4 some results of the MEVT will be discussed and finally the backtesting results will be reported in Section 5.5.

## 5.3 Vine Copula results

To illustrate the performance of the Vine Copula at a certain point in time, the dataset prior to 2007-07-01 containing 1500 observations will be used. These observations will be filtered by using the marginal modeling as reported earlier. Once the observations have been filtered a pair-scatter plot and a pair-contour plot is constructed, which is shown in the Appendix C in Figure 22. It can easily be observed that there is a strong relationship between the indices in Europe by looking at the Kendall's  $\tau$ . Furthermore the Asian indices show among each other quite a strong relationship, while the relationship with the indices in Europe and the S&P500 is rather limited. The S&P500 shows a stronger relationship with the European indices than with the Asian ones. These results can also be seen in the selected Vine structure in the Appendix C in Figure 23. In the first tree the very strong relationships between the AEX, the CAC40 and the DAX estimated by using bivariate t copulas, which indicate the presences of fat tails and tail dependence. The CAC40 and the SMI are also linked, which is expected due to the geographical argument. The S&P500 is linked to the SMI, while the Hang Seng and the NIKKEI225 are linked to each other as well. The other trees are hard to interpret and the dependence in these trees decrease tremendously, which can be observed in the Appendix C in Figure 24.

## 5.4 MEVT results

In this section we will discuss some results of the joint distribution using MEVT. For illustration purposes we choose to report the values of  $a^-$  for the AEX, CAC40, DAX, Hang Seng and the S&P500. These results are shown in a smoothed path (spline) over time. The values have been estimated in a rolling window setting containing 1500 observations and a tail of 5% starting at 2007-07-01 until 2009-07-01. Since we expect that the Financial Crisis is interesting for the tail behavior we choose to report these results, which can be found in the Appendix C in Figures 25-28. We show two different lines in each subplot because the effects are not required to be symmetric. In other words the value of  $a_{j|i}^-$  may differ from  $a_{i|j}^-$ . The black line indicates the value  $a_i^-$  given the fact that  $j$  is a negative extreme. For example the Figure 25 corresponds with with the values of  $a^-$  for all the indices except the AEX conditioned the fact that the AEX is a negative extreme. The estimated values of  $a^-$  are particularly

interesting since they tell us something about the lower asymptotic (tail) dependence. From section 4.7 it follows that the following holds:

1. Positive asymptotic dependence if  $0 < a_{j|i}^\dagger < 1$  or **both**  $a_{j|i}^\dagger = 0$  **and**  $b_{j|i}^\dagger > 0$ .
2. Negative asymptotic dependence if  $-1 < a_{j|i}^\dagger < 0$ .
3. Close to asymptotic independence if **both**  $a_{j|i}^\dagger = 0$  **and**  $b_{j|i}^\dagger \leq 0$

If we look at figure 25 in the Appendix C we see that during the crisis the value of  $a_{j|AEX}^-$ ,  $\forall j \in \{\text{CAC40, DAX, SMI}\}$  decreases until mid 2008. This can be explained by the fact that the data of the dotcom bubble period leaves the in sample. When the Financial Crisis hits its peak in the mid 2008, we observe a shift in this pattern. The tail dependence starts to increase strongly for the CAC40, DAX, SMI and the AEX. This is expected given the domino-like collapse of the financial markets during the Financial Crisis in 2008. The value of  $a_{j|AEX}^-$ ,  $\forall j \in \{\text{Hang Seng, NIKKEI225}\}$  is very close to zero, indicating asymptotic independence in the lower tails. This can be explained by the fact that these markets are not integrated as much as for example other European markets and the US market. The asymptotic dependence between the AEX and the S&P500 seems to be positive and more stable over time. Figure 27 shows that the influence in terms of lower asymptotic dependence of the Hang Seng index on the other indices except the NIKKEI is rather limited. The influence of the Hang Seng index on the NIKKEI225 index is quite extensive. We think this is a reasonable result, since both markets are financially strongly integrated due to mainly the geographic locations of both markets. There seems to be a negative relationship forming over time between many indices and the Hang Seng index. The S&P500 and the SMI are exceptions, they seem to behave asymptotic independent in the lower tails with the Hang Seng index. The Hang Seng index contains lots of Chinese blue chip companies. China was until 2015 the fastest growing economy, with extreme growth rates. Therefore it is no surprise that the Hang Seng index behaves quite different from the other indices. In Figure 28 we observe that the influence of the S&P500 on the European indices is positive and increasing slightly over time. This can be explained by the fact that the American economy is a strong and persistent driver for the European economies due to strong market integration.

## 5.5 Backtesting Results

In this section we use three different dependence models a moving window setting with 1500 observations that is re-estimated over time. The dependence models we use are Vine Copulas and a model based on MEVT. The t copula is used as the benchmark model. We use the backtests to evaluate the performance of the dependence models. We choose to use a large moving window, since many observations are required for putting this type of MEVT into practice. The MEVT requires modeling assumption from Heffernan & Tawn (2004) (see Section 4.7) which only holds if the observations in the tail are extreme, otherwise a bias is introduced.

The backtesting is split up into two difference time periods. One time period includes the Financial Crisis (2007-07-01 until 2009-07-01) in the out sample. The other time period excludes the Financial Crisis in both the in and out sample (2015-01-01 until 2017-01-01). We estimate both the one-day-ahead VaR/ES and the ten-day-ahead VaR/ES. We use two different portfolio constructions: an equally-weighted portfolio and a minimum variance portfolio. Both portfolios imply different weights to the assets. If the weight vector is multiplied by real returns or the simulated returns, the portfolio returns can be constructed.

In Tables 12 and 13 we report the lower breaches, using an equally-weighted portfolio, of the estimated one- and ten-day-ahead VaR for the first dataset. Moreover we report in these tables the results of the conditional and unconditional coverage Kupiec's and the results of the ES zero mean test.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	38,9	<b>47</b>	51	53
5%	19,45	<b>26</b>	29	<b>24</b>
2%	7,78	10	14	9
1%	3,89	9	10	7
0,50%	1,945	3	9	1

Table 6: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . The column *Expected* indicates the expected number of breaches given *alpha*. This corresponds to multiplying the value  $\alpha$  by the total number of observations. Portfolios are constructed using equal weights. This set contains 389 observations. The gray digits indicate that the model does not capture the number of breaches correctly according to the unconditional coverage test using a 5% significance level. The **bold** numbers indicate that both the conditional and the unconditional coverage tests support the assumption of correct number of lower breaches. The *italic* number indicates that the ES is specified correctly using the zero mean test.

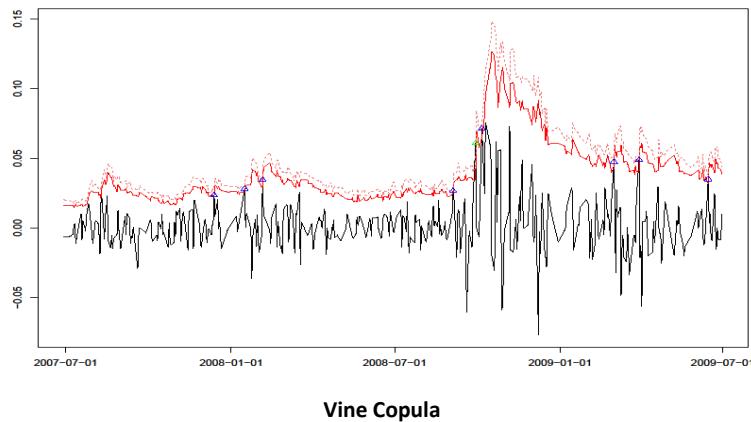
<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	38,9	71	84	72
5%	19,45	31	45	32
2%	7,78	14	21	12
1%	3,89	10	12	10
0,50%	1,945	7	10	7

Table 7: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights.

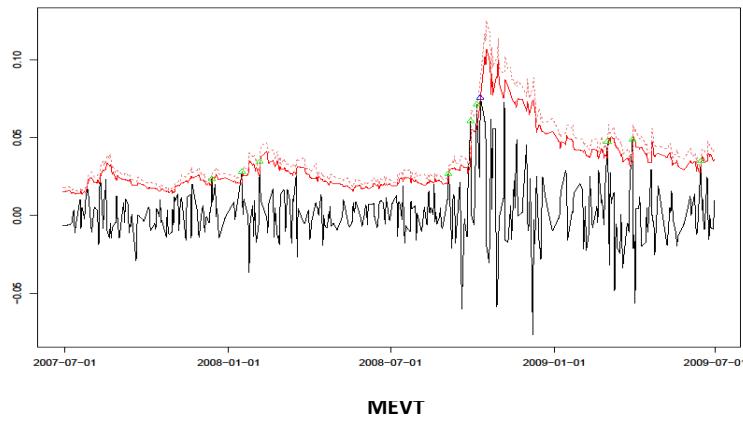
When we compare the expected number of breaches with the realized breaches from each method, we can tell something about the performance of each method. From Table 12 we observe that all the methods seem to underestimate the VaR and the ES, since the number of breaches is higher than expected number of breaches given *alpha*. This VaR underestimation is however no surprise. After all the out sample contains many extreme observations due to the Financial Crisis. We observe an underestimation of the ten-day-ahead VaR and ES forecasts as well (Table 13). This underestimation is however more severe than the underestimation of the one-day-ahead forecast. We can explain this because we forecast ten-day-ahead cumulative returns in Table 13. If for example a very extreme loss is exactly ten days in front of the last observation in the in sample, then it takes ten days before the information about this extreme loss enters the in sample. Furthermore we know from the stylized fact from McNeil et al. (2005) that extremes occur often in clusters. This may cause several consecutive breaches. In the following Figures 4 and 5 both the daily and cumulative ten day log losses are shown over 2007-07-01 until 2009-07-01 (389 observations) along with their 99% VaR and the ES estimations and the breaches. In Figure 5 we observe the clustering of breaches for all the methods when forecasting the ten-day-ahead 99% VaR and ES. If we compare the number of breaches of the one-day-ahead VaR and ES estimations (Table 12) among the different methods, we observe that the t copula outperforms the other two methods for generally all the values of *alpha*. It follows that the unconditional coverage test indicates that the t copula method significantly produces the correct number of breaches for  $\alpha = 5\%, 2\%, 1\%$  and  $0.5\%$ . The number breaches of the method based on Vine Copula is slightly higher but still very close to the number of breaches of the method when a t copula is used. This can be explained by the fact that many bivariate relationships in the first tree, which is the most important one in the Vine structure, are estimated using bivariate t copulas (for example see Figures 22 and 23). The major difference is that for the Vine Copula method in later trees many bivariate conditional couples are estimated by an independence copula. This reduces the computation time, but it also reduces the dependence. Therefore, this method produces in this case tails that are less fat than the t copula. The fat tails of the t copula are advantageous when estimating the VaR or ES during a period of financial distress. However, generally speaking, the in sample fit of the Vine copula is higher in terms of likelihood than the in sample fit of the t copulas (Aas et al. 2009).

The performance of the method based on MEVT in terms of breaches is the worst. The number of breaches for  $\alpha = 1\%$  however, which is in the main point of interest given Basel II, is very close to the number of breaches of the other methods. For none of the  $\alpha$ 's the unconditional coverage test indicate the correct number of breaches significantly. The method based on MEVT seem to have some trouble with catching up. Again this can be explained due to overfitting of the t copula and the Vine Copula, that is quite similar to the t copula. Especially during the peak of the financial crisis the method based on MEVT seem to be not very reactive. This may be caused by the fact that the length of the window is quite long. If we look at Figures 4 and 5 we indeed observe that the 99% quantile of the portfolio losses is the highest for the method based on the t copula. The Vine Copula has a lower 99% quantile estimation, while the method based on MEVT has the lowest.

Equal weights portfolio daily losses 2007-07-01 until 2009-07-01



**Vine Copula**



**MEVT**

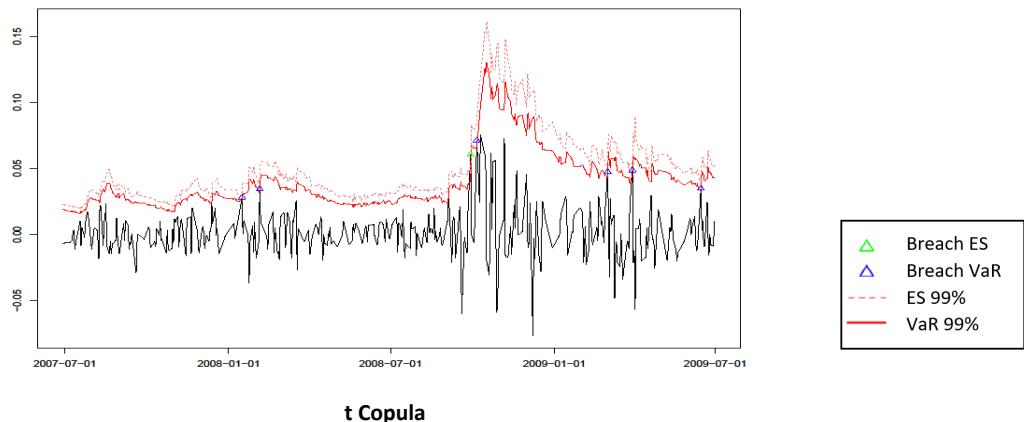


Figure 4: This figure contains the daily log losses along with the estimated one-day-ahead 99% VaR and ES and the breaches.

Equal weights portfolio ten day cummulative losses 2007-07-01 until 2009-07-01

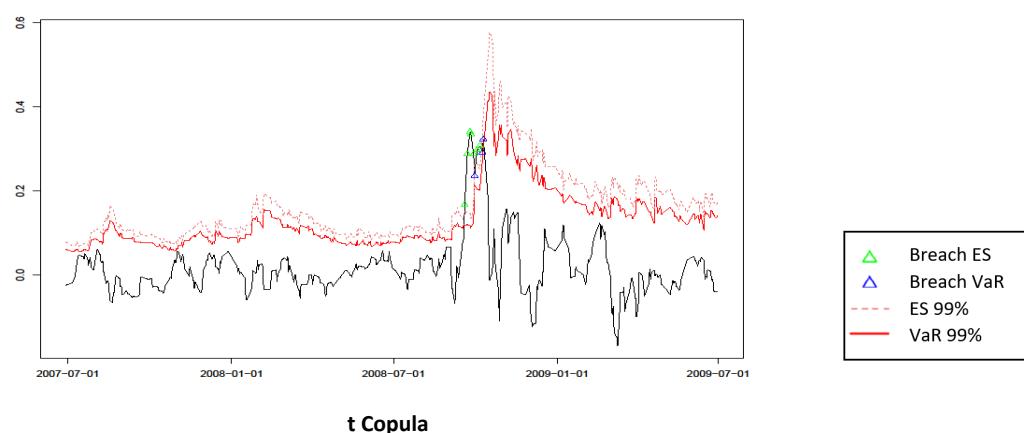
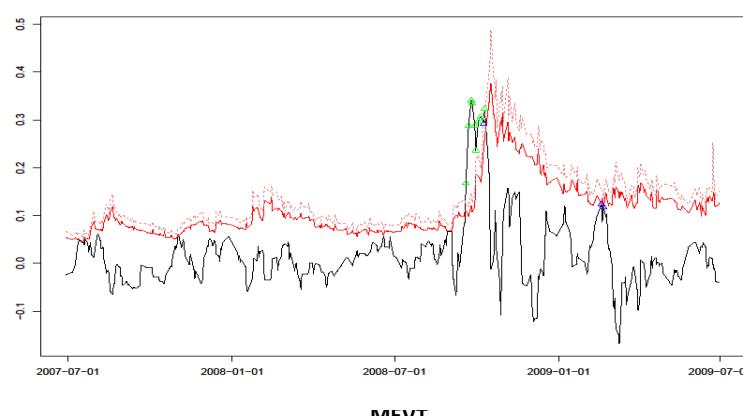
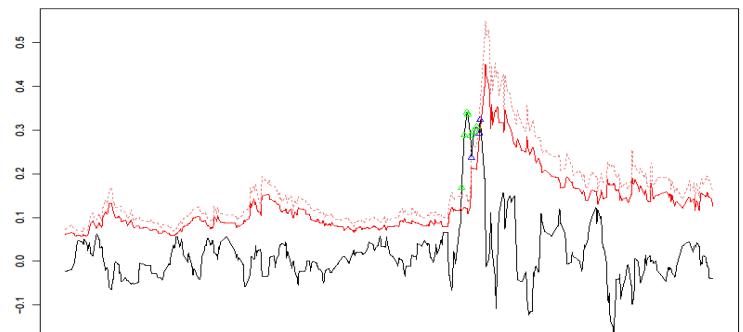


Figure 5: This figure contains the cumulative ten day log losses along with the estimated ten-day-ahead 99% VaR and ES and the breaches.

Since portfolio managers often want to diversify and limit the risk, we believe it is important to analyze the performance of the dependence models in a less risky portfolio. We therefore choose to construct a minimum variance portfolio that is not rebalanced for simplicity and have a short position constraint <sup>16</sup>. In Tables 14 and 15 in Appendix B the breaches are shown for this minimum variance portfolio. The weights are reported in the tables as well. It follows that there is no weight put on the AEX, CAC40 and the DAX, which are strongly correlated to each other. Most of the weight is put on the S&P500 (41%), the Hang Seng (28%) and the SMI (22%). The number of breaches shrinks quite a bit regarding the equally weighted portfolio. The tests show that the one-day-ahead VaR and the ES of 2%, 1% and 0.5% for all of the all three methods is well defined. The ten-day-ahead VaR and ES are according to the tests not correctly specified.

We can conclude from both the graphs and the tables that during financial distress we would prefer to use t copula, whose performance is slightly better than the performance of the Vine copula. The MEVT seems to have some trouble catching up to the other two methods.

To analyze the performance of our dependence model during a stable financial period we choose to use 371 observations from 2015-01-01 until 2017-01-01 with the same moving window length. We choose to use these dates since we do not want our in sample to be interfered by the extreme events during Financial Crisis. In tables 16 and 17 the breaches are shown for both the one-day- and cumulative ten-day-ahead VaR.

alpha	Expected	Vine Copula	MEVT	t Copula
10%	39,1	22	<b>33</b>	26
5%	19,55	<b>12</b>	<b>15</b>	<b>14</b>
2%	7,82	5	6	5
1%	3,91	3	3	3
0,50%	1,955	2	3	2

Table 8: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 391 observations.

alpha	Expected	Vine Copula	MEVT	t Copula
10%	39,1	20	28	21
5%	19,55	10	<b>14</b>	11
2%	7,82	5	7	5
1%	3,91	3	5	3
0,50%	1,955	3	3	3

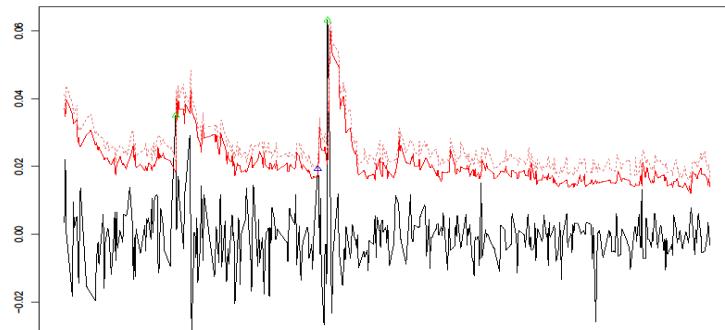
Table 9: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 391 observations.

We observe from table 16 that both the Vine Copula and the t copula produce fewer breaches than expected. This overestimation of the VaR and ES may be explained by the fact that both methods have quite fat tails. The upper and lower tail dependence of the t copula is similar. Vine copulas often use bivariate t copulas in the first trees. Therefore, these copulas show some symmetry as well. This is not the case for the method based on MEVT. Since this period is known for its recovery after the Financial Crisis, which can be observed in Figures 1 and 2, the positive shocks in returns affect the estimation of the lower quantile when using t copulas or Vine Copulas. This will lead to an overestimation of the lower tail. This can also be observed in both tables 16 and 17. The method based on MEVT outperforms the other two methods based on the breaches. In Figures 6 and 7 the daily and the cumulative ten-day log losses are shown over time along with the 99% VaR and ES estimation and the breaches. In these figures we see that both the Vine Copula and the t copula produce higher VaR and ES estimates than the method based on MEVT. The same results can be drawn from the breaches of the minimum

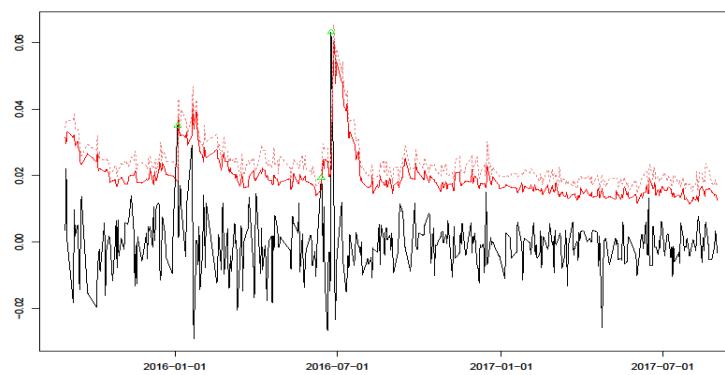
<sup>16</sup>This means that short positions are not allowed

variance portfolio reported in Tables 18 and 19 in the Appendix B. We can conclude that in more stable financial periods we prefer to use methods based on MEVT above the Vine Copula and the t copula

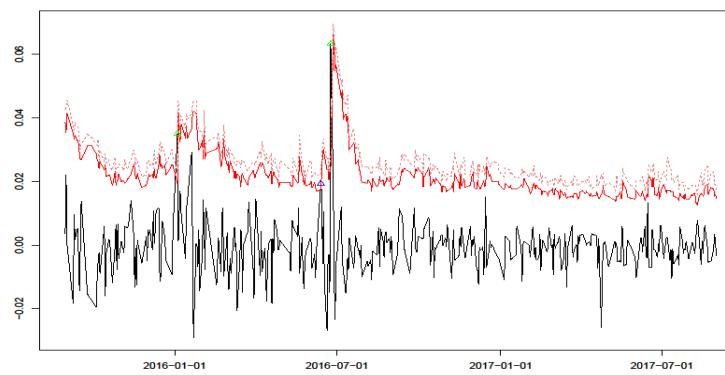
### Equal weights portfolio daily losses 2015-01-01 until 2017-01-01



**Vine Copula**



**MEVT**

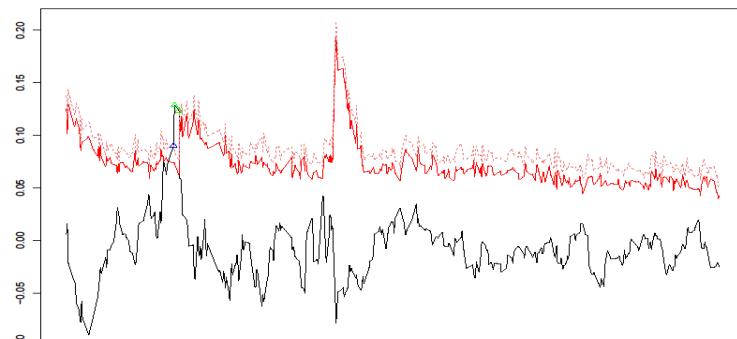


▲	Breach ES
△	Breach VaR
- - -	ES 99%
—	VaR 99%

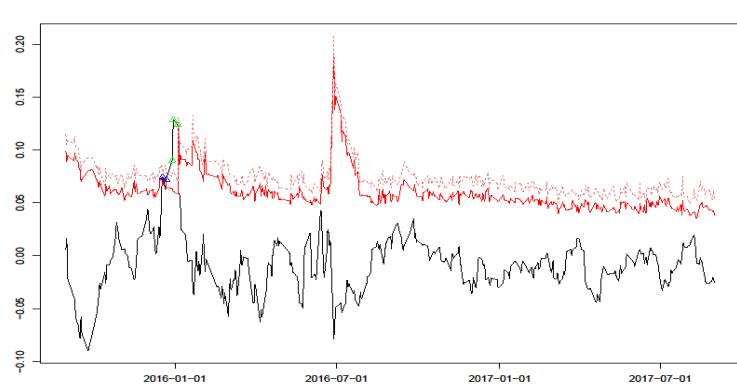
**t Copula**

Figure 6: This figure contains the daily log losses along with the estimated one-day-ahead 99% VaR and ES and the breaches.

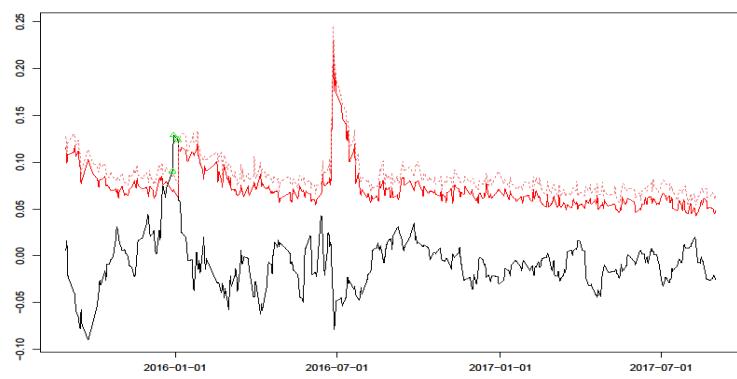
Equal weights portfolio ten day cummulative losses 2015-01-01 until 2017-01-01



Vine Copula



MEVT



- ▲ Breach ES
- ▲ Breach VaR
- - ES 99%
- VaR 99%

t Copula

Figure 7: This figure contains the cumulative ten day log losses along with the estimated ten-day-ahead 99% VaR and ES and the breaches.

## 6 Conclusion

In this research we forecast both the one- and ten-day-ahead Value-at-Risk (VaR) and Expected Shortfall (ES) for a portfolio using two different dependence models: Vine Copula and a method based on Multivariate Extreme Value Theory (MEVT). We use two steps to do so. Firstly we model the marginals and secondly we model the dependence structure. The marginal modeling consists of the two-step method. In the first step we filter the marginals by a Markov-Switching variant of the ARMA-GARCH model with an asymmetric component (GJR). In the second step we model the filtered returns semi-parametrically using extreme value theory. In the end we compare the performance of our dependence models with a t copula by using backtesting with conditional and unconditional coverage tests for the VaR and the zero mean test for the ES. We use two different samples for backtesting the VaR and ES of a portfolio consisting seven different major world indices in a rolling window setting containing 1500 observations. Firstly we evaluate the performance of our models during the financial crisis (2007-07-01 until 2009-07-01) and secondly during a stable financial period (2015-01-01 until 2017-01-01).

It follows that the MEVT underestimates the VaR during the financial crisis. The fat tails of the t copula and the Vine copula seem to capture the downward breaches better than the model based on MEVT. Generally, all the models underestimate the ten-day-ahead VaR during the crisis and too many breaches occur. The one-day-ahead VaR estimates are much more precise in terms of breaches for all the models than the ten-day-ahead VaR estimates. Especially the t copula is able of significantly producing the correct number of breaches for  $\text{VaR}_{95\%}$ ,  $\text{VaR}_{98\%}$ ,  $\text{VaR}_{99\%}$  and  $\text{VaR}_{99.5\%}$ .

In the stable financial period both the Vine copula and the t copula seem to have tails that are too fat. These methods tend to overestimate the VaR. The MEVT however, captures the number of breaches more correctly. The ten-day-ahead VaR estimations of this dataset are much more precise for all of the methods than during the financial crisis.

We can conclude that the VaR estimation performance of the Vine copula regarding the t copula is quite comparable. Both methods produce heavy tails which is beneficial during periods of financial distress, but lead to overestimation during stable financial periods. Unlike the Vine Copula and the t copula, the model based on MEVT slightly underestimates the VaR and the ES during financial crisis. However, this method seems to outperform the other two during stable periods.

We see that the performance of the dependence methods varies over time. In order to overcome this issue we think it is beneficial to make the dependence model more robust to regime changes. It may be beneficial to create a dependence structure that accommodates different regimes. Therefore, we believe that this is a good suggestion for further research.

## References

Aas, K. (2013), 'Pair-copula constructions: even more flexible than copulas', *Norwegian Computing Center* .

Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009), 'Pair-copula constructions of multiple dependence', *Insurance: Mathematics and Economics* **44**(2), 182–198.

Artzner, P., Delbean, F. & David Heath, P. (1999), 'Coherent measures of risk', *Mathematical Finance* **9**, 203–228.

Bauwens, L., Preminger, A. & Rombouts, J. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* p. 357.

Bauwens, L., Preminger, A. & Rombouts, J. (2010), 'Theory and inference for a markov switching garch model', *The Econometrics Journal* p. 218.

Bedford, T. & Cooke, R. (2002), 'Vines: A new graphical model for dependent random variables', *The annals of Statistics* **30**(4), 1031–1068.

Berg, D. (2009), 'Copula goodness-of-t testing: An overview and power comparison', *The European Journal of Finance* p. 675.

Bouyé, E., Roncalli, T., Durrleman, V., Nikeghbali, A. & Riboulet, G. (2000), 'Copulas for finance: A reading guide and some applications', *Groupe de Recherche Opérationnelle, Bercy-Expo* .

Brockwell, P. & Davis, R. (1991), 'Time series: Theory and methods', *Springer Series in Statistics* .

Chava, S., Stefanescu, C. & Turnbull, S. (2008), 'Modeling the loss distribution', *Working Paper* .

Christoffersen, B. P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**, 841–862.

Clarke, K. (2007), 'A simple distribution-free test for nonnested model selection', *Political Analysis* **15**, 347–363.

Cormen, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. (2001), 'Introduction to algorithms', *Cambridge: the MIT Press* **2**.

Czado, C., Jaworski, P., Durante, F., Härdle, W. & Rychlik, T. (2010), *The Asymptotic Theory of Extreme Order Statistics*, Springer, Berlin, Heidelberg.

Danielsson, J., de Haan, L., Ergun, L. & de Vries C (2016), 'Tail index estimation: Quantile driven threshold selection', Available at SSRN: <https://ssrn.com/abstract=2717478> .

Dißmann, J. (2010), Statistical Inference for Regular Vines and Application, PhD thesis, Technische Universität München.

Efron, B. & Tibshirani, R. (1994), *An Introduction to the Bootstrap*, Chapman and Hall.

Embrechts, P., Klüppelberg, C. & T., M. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer.

Embrechts, P., Lindskog, F. & McNeil, A. (2003), 'Modelling dependence with copulas and applications to risk management', *Handbook of Heavy Tailed Distributions in Finance* .

Embrechts, P., McNeil, A. & Straumann, D. (1999), 'Correlation: pitfalls and alternatives', *Risk* **12**, 69–71.

Embrechts, P., Resnick, S. & Samorodnitsky, G. (1999), 'Living on the edge', *RISK magazine* **1**, 96–100.

Fredricks, G. & Nelsen, R. (2007), 'On the relationship between spearmans rho and kendalls tau for pairs of continuous random variables', *Journal of statistical planning and inference* p. 2143.

Genest, C. & Rémillard, B. (2007), 'Everything you need to know about copula modeling but were afraid to ask', *Journal of Hydrologic Engineering* **12**, 347–368.

Glosten, L., Jagannathan, R. & Runkle, D. (1993), 'On the relation between the expected value and the volatility of the nominal excess return on stocks', *Journal of Finance* p. 1779.

Green, P. & Richardson, S. (2003), *Highly Structured Stochastic Systems*, Oxford: Oxford University Press.

Gruber, L. & Czado, C. (2015), 'Sequential bayesian model selection of regular vine copulas', *Bayesian Analysis* **10**(4), 937–963.

Heffernan, J. & Tawn, J. (2004), 'A conditional approach for multivariate extreme values', *Journal of the Royal Statistical Society* **3**, 497–546.

Hilal, S., Poon, S. & Tawn, J. (2011), 'Hedging the black swan: Conditional heteroskedasticity and tail dependence in S&P500 and VIX', *Journal of banking & Finance* **35**, 2374–2387.

Hilal, S., Poon, S. & Tawn, J. (2014), 'Portfolio risk assessment using multivariate extreme value methods', *Extremes* **17**(4), 531–556.

Hsu, C., Huang, C. & Chiou, W. (2012), 'Effectiveness of copula-extreme value theory in estimating Value-at-Risk: empirical evidence from asian emerging markets', *Review of Quantitative Finance and Accounting* p. 447.

Joe, H. (1996), 'Families of m-variate distributions with given margins and  $m(m-1)/2$  bivariate dependence parameters', *Lecture Notes-Monograph Series* **28**.

Joe, H. (1997), 'Multivariate models and dependence concepts', *Monographs on Statistics and Applied Probability* (73).

Kupiec, P. (1995), 'Techniques for verifying the accuracy of risk management models', *Journal of Derivatives* p. 73.

Kurowicka, D. (2010), 'Optimal truncation of vines', *Dependence Modeling: Handbook on Vine Copulae* .

Kurowicka, D. & Cooke, R. (2006), 'Uncertainty analysis with high dimensional dependence modeling', *John Wiley and Sons* .

Ledford, A. & Tawn, J. (1996), 'Statistics for near independence in multivariate extreme values', *Biometrika* **83**, 169–187.

Liu, Y. & Luger, R. (2009), 'Efficient estimation of copula-garch models', *Computational Statistics & Data Analysis* p. 2284.

Ljung, G. & Box, G. (1978), 'On a measure of a lack of fit in time series models', *Biometrika* **65**, 297–303.

Longin, F. (1997), 'From Value-at-Risk to stress testing, the extreme value approach', *CERESSEC* .

Mabrouk, S. & Chaker, A. (2011), 'Garch class models estimations and Value-at-Risk analysis for exchange rate', *International Jounal of Monetary Economics and Finance* p. 254.

McNeil, A. (1998), 'Calculating quantile risk measures for financial return series using extreme value theory', *ETH Zürich* .

McNeil, A. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**, 271–300.

McNeil, J. A., Frey, R. & Paul, E. (2005), *Quantitative Risk Management*, Princeton University Press, Princeton and Oxford.

Mikosch, T. & Starica, C. (2004), 'Nonstationarities in financial time series, the long-range dependence, and the igarch effects', *Review of Economics and Statistics* p. 378.

Nelsen, R. (1999), *An Introduction to Copulas*, Springer, New York.

Nyström, K. & Skoglund, J. (2002), 'A framework for scenario-based risk management'.

Poon, S., Rockinger, M. & Tawn, J. (2004), 'Extreme-value dependence in financial markets: diagnostics, models and financial implications', *Review of Financial Studies* **17**, 581–610.

Prim, R. (1957), 'Shortest connection networks and some generalizations', *Bell System Technical Journal* **36**, 1389–1401.

Rootzen, H. & Tajvidi, N. (2000), 'Extreme value statistics and wind storm losses: a case study', *Extremes and Integrated Risk Management* .

Savu, C. & Trede, M. (2006), 'Hierarchical archimedean copulas', *International Conference on High Frequency Finance, Konstanz, Germany* .

Schepsmeier, U. & Czado, C. (2015), 'Dependence modelling with regular vine copula models: a case study for car crash simulation data', *Royal Statistical Society* .

Schweizer, B. & Sklar, A. (1983), *Probabilistic metric spaces*, Courier Corporation.

Sklar (1959), 'Fonctions de répartition à n dimensions et leurs marges', *Publications de l'Institut de Statistique de L'Université de Paris* **8**, 229–231.

Stöber, J., Joe, H. & Czado, C. (2013), 'Simplified pair copula constructions - limitations and extensions', *Journal of Multivariate Analysis* **119**, 101–118.

Tang, W. & Gau, Y. (2004), 'Forecasting Value-at-Risk using the markov-switching arch model'.

Vuong, Q. (1989), 'Likelihood ratio tests for model selection and non-nested hypotheses', *Econometrica* **57**, 307–333.

Zhang, Z. & Huang, J. (2006), 'Extremal financial risk model and portfolio evaluation', *Comput. Stat. Data Anal* **51**, 2313–2338.

# Appendix

## A Derivations

### A.1 Hamilton Filter

Let's consider a two state Markov-Switching model where the innovations are assumed to be t-distributed. Let  $S_t$  denote the unobserved regimes, which can take values 1 or 2. Then the transition matrix  $P$  can be described as:

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}, \quad \text{with}$$

$$\begin{aligned} p_{11} &= P(S_t = 1 | S_{t-1} = 1) \\ p_{21} &= P(S_t = 2 | S_{t-1} = 1) = 1 - P(S_t = 1 | S_{t-1} = 1) = 1 - p_{11} \\ p_{22} &= P(S_t = 2 | S_{t-1} = 2) \\ p_{12} &= P(S_t = 1 | S_{t-1} = 2) = 1 - P(S_t = 2 | S_{t-1} = 2) = 1 - p_{22} \end{aligned}$$

The Hamilton filter consists of iteratively repeating two consecutive steps. Firstly the model is predicted and secondly the model is updated.

#### A.1.1 Prediction Step

$$P(S_t = s_t | I_{t-1}) = \sum_{i=1}^2 P(S_t = s_t | S_{t-1} = i) P(S_{t-1} = i | I_{t-1}),$$

where  $I_t$  denotes the information set available at time  $t$  (note that  $I_t = I_{t-1} \cup X_t$ ) and  $P(S_{t-1} = s_{t-1} | I_{t-1})$  is the result of updating step in the previous iteration.

$$f(X_t | I_{t-1}) = \sum_{i=1}^2 f(X_t, S_t = i | I_{t-1}) = \sum_{i=1}^2 f(X_t | S_t = i, I_{t-1}) P(S_t = i | I_{t-1})$$

Using the equations both the probability of being in a certain state one period ahead is evaluated along with the one day ahead prediction of the distribution of  $X_t$ .

#### A.1.2 Updating Step

$$P(S_t = s_t | I_t) = \frac{f(X_t, S_t = s_t | I_{t-1})}{f(X_t | I_{t-1})} \quad (48)$$

The Hamilton filter requires initialization. It is often initialized using the unconditional probabilities of being in certain states. These probabilities can be found by finding vector  $v$  such that:

$$Pv = v$$

In other words this corresponds to the vector which remains the same when it is multiplied by the transition matrix. Using this equation, it turns out that this vector corresponds to the eigenvector. If we consider two different Markov states then:

$$Pv = \begin{pmatrix} p_{11} & (1 - p_{22}) \\ (1 - p_{11}) & p_{22} \end{pmatrix} v = v, \quad \text{where}$$

$$v = \begin{pmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{pmatrix}$$

where  $\frac{1-p_{22}}{2-p_{11}-p_{22}}$  denotes the unconditional probability of being in state 1 and  $\frac{1-p_{11}}{2-p_{11}-p_{22}}$  denotes the unconditional probability of being in state 2. We start the filter with  $P(S_0 = 1) = \frac{1-p_{22}}{2-p_{11}-p_{22}}$  and  $P(S_0 = 2) = \frac{1-p_{11}}{2-p_{11}-p_{22}}$ .

### A.1.3 Estimation of Hamilton Filter

We can now obtain the Maximum Likelihood function  $\mathcal{L}$ :

$$\mathcal{L}(X_T) = \prod_{t=1}^T f(X_t | I_{t-1})$$

## A.2 Marginal GPD

In this section the marginal GPD is derived. The aim is to construct the following:

$$\widehat{F}_{X_i}(x) = \begin{cases} \widetilde{F}_{X_i}(u_X^-)[1 + \xi^-(u_X^- - x)/\phi^-]_+^{-1/\xi^-} & \text{if } x < u_X^-, \\ \widetilde{F}_{X_i}(x) & \text{if } x \in [u_X^-, u_X^+], \\ 1 - \widetilde{F}_{X_i}(u_X^+)[1 + \xi^+(x - u_X^+)/\phi^+]_+^{-1/\xi^+} & \text{if } x > u_X^+ \end{cases} \quad (49)$$

It does not require any explanation to show that  $\widehat{F}_{X_i}(x) = \widetilde{F}_{X_i}(x)$  when  $x \in [u_X^-, u_X^+]$ . Next the third line of the equation is considered. We know that  $\widetilde{F}_{X_i}(u_X^+) = 1 - \widetilde{F}_{X_i}(u_X^+)$ . The main interest in this case is to model the probability that the random variable  $X$  is less than a certain constant  $x$  given the case that  $x > u$  ( $x$  exceeds a certain threshold  $u$  and therefore considered to be in the tail)<sup>17</sup>:

$$\begin{aligned} Pr(X \leq x | x > u) &= F(x | x > u) = 1 - \overline{F}(x | x > u) \\ \overline{F}(x | x > u) &= Pr(X > x | x > u) \frac{1}{Pr(X > x | x > u)} Pr(X > x | x > u) \\ &= Pr(X > u | x > u) \frac{Pr(X > x, X > u)}{Pr(X > u)} \\ &= \overline{F}(u) Pr(X - u > x - u | X > u) \\ Pr(X \leq x | x > u) &= 1 - \overline{F}(u) [1 - Pr(X - u < x - u | X > u)], \end{aligned}$$

where  $\overline{F}(u)$  can be replaced by the empirical estimate. Furthermore it is known that  $Pr(X - u > x - u | X > u)$  follows a GPD. If the indices are placed back the following expression can be obtained:

$$\widehat{F}_{X_i}(x) = 1 - \widetilde{F}_{X_i}(u_X^+)[1 + \xi^+(x - u_X^+)/\phi^+]_+^{-1/\xi^+} \quad \text{if } x > u_X^+$$

In the same way the first equation can be proved to be correct.

## A.3 Five dimensional C-Vine and D-Vine

$$\begin{aligned} f(x_1, x_2, x_3, x_4, x_5) &= f(x_1) \cdot f(x_2) \cdot f(x_3) \cdot f(x_4) \cdot f(x_5) \\ &\quad \cdot c_{12}(F(x_1), F(x_2)) \cdot c_{13}(F(x_1), F(x_3)) \cdot c_{14}(F(x_1), F(x_4)) \cdot c_{15}(F(x_1), F(x_5)) \\ &\quad \cdot c_{23|1}(F(x_2|x_1), F(x_3|x_1)) \cdot c_{24|1}(F(x_2|x_1), F(x_4|x_1)) \cdot c_{25|1}(F(x_2|x_1), F(x_5|x_1)) \\ &\quad \cdot c_{34|12}(F(x_3|x_1, x_2), F(x_4|x_1, x_2)) \cdot c_{35|12}(F(x_3|x_1, x_2), F(x_5|x_1, x_2)) \\ &\quad \cdot c_{45|123}(F(x_4|x_1, x_2, x_3), F(x_5|x_1, x_2, x_3)) \end{aligned} \quad (50)$$

---

<sup>17</sup>For practical issues the indices are dropped for the derivation and later on added.

## B Tables

ARMA(i,j)-MS(2)-GARCH(k,l)-GJR(m)					
Marginal	i	j	k	l	m
AEX	0	0	1	1	1
CAC40	0	0	1	1	1
DAX	0	0	1	1	1
Hang Seng	0	0	1	1	1
NIKKEI225	0	0	1	1	1
SMI	0	0	1	1	1
S\&P500	0	0	1	1	1

Start: 5-1-2000
End: 28-6-2007
Number of Observations: 1500

Table 10: This table contains the number of lags selected by the BIC criterion for the indices dataset. The number of lags is calibrated on a time period starting in 5-1-2000 until 28-6-2007 and kept fixed over the  $n$ -day ahead forecasts.

	Ljung-Box (PValue)	Augmented DF (PValue)
AEX	0,886	<0,01
CAC40	0,583	<0,01
DAX	0,732	<0,01
Hang Seng	0,053	<0,01
NIKKEI225	0,993	<0,01
SMI	0,930	<0,01
S\&P500	0,219	<0,01

Start: 5-1-2000
End: 28-6-2007
Observations: 1500

Table 11: This table shows the P-values of the Ljung-Box and the Augmented Dickey-Fuller test with ten lags. Under the null hypothesis of the Ljung-Box test is assumed that the data is randomly distributed. Under the null hypothesis of the Augmented Dickey-Fuller test the data is assumed to be non-stationary. Both tests are conducted using the Indices dataset from 5-1-2001 to 28-6-2007. The **bold** numbers represent the case when the P-values of the Ljung-Box test are below 0.05% or when the P-values of the Augmented Dickey-Fuller test are above 0.05%.

alpha	Expected	Vine Copula	MEVT	t Copula
10%	38,9	<b>47</b>	51	53
5%	19,45	<b>26</b>	29	<b>24</b>
2%	7,78	10	14	9
1%	3,89	9	10	7
0,50%	1,945	3	9	1

Table 12: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 389 observations. The gray digits indicate that the model does not capture the number of breaches correctly according to the unconditional coverage test using a 5% significance level. The **bold** numbers indicate that both the conditional and the unconditional coverage tests support the assumption of correct number of lower breaches. The *italic* number indicates that the ES is specified correctly using the zero mean test.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	38,9	71	84	72
5%	19,45	31	45	32
2%	7,78	14	21	12
1%	3,89	10	12	10
0,50%	1,945	7	10	7

Table 13: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 389 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	38,9	49	53	55
5%	19,45	24	29	25
2%	7,78	9	10	6
1%	3,89	5	8	3
0,50%	1,945	3	4	2

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
Weights	0%	0%	0%	28%	9%	22%	41%

Table 14: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . Portfolios are constructed using a minimum variance portfolio, constructed using 1000 observation prior to 2007-07-01 (no rebalancing). This set contains 389 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	38,9	58	73	58
5%	19,45	36	41	37
2%	7,78	16	23	20
1%	3,89	11	14	10
0,50%	1,945	8	11	9

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
WEIGHTS	0%	0%	0%	28%	9%	22%	41%

Table 15: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2007-07-1 until 2009-07-01 for several different values of  $\alpha$ . Portfolios are constructed using a minimum variance portfolio, constructed using 1000 observation prior to 2007-07-01 (no rebalancing). This set contains 389 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	39,1	22	33	26
5%	19,55	12	15	14
2%	7,82	5	6	5
1%	3,91	3	3	3
0,50%	1,955	2	3	2

Table 16: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 391 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	39,1	20	28	21
5%	19,55	10	14	11
2%	7,82	5	7	5
1%	3,91	3	5	3
0,50%	1,955	3	3	3

Table 17: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using equal weights. This set contains 391 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	39,1	18	23	23
5%	19,55	11	12	11
2%	7,82	3	4	3
1%	3,91	3	3	3
0,50%	1,955	3	3	2

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
WEIGHTS	0%	0%	0%	7%	32%	25%	36%

Table 18: This table contains the number of one-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using a minimum variance portfolio, constructed using 1000 observation prior to 2015-01-01 (no rebalancing). This set contains 391 observations.

<i>alpha</i>	Expected	Vine Copula	MEVT	t Copula
10%	39,1	15	24	18
5%	19,55	11	12	11
2%	7,82	5	7	4
1%	3,91	3	4	2
0,50%	1,955	2	3	2

	AEX	CAC40	DAX	Hang Seng	NIKKEI225	SMI	S&P500
WEIGHTS	0%	0%	0%	7%	32%	25%	36%

Table 19: This table contains the number of ten-day-ahead  $\text{VaR}_\alpha$  lower breaches for all of the three dependency models of the indices dataset from 2015-01-1 until 2017-01-01 for several different values of  $\alpha$ . Portfolios are constructed using a minimum variance portfolio, constructed using 1000 observation prior to 2015-01-01 (no rebalancing). This set contains 391 observations.

## C Figures

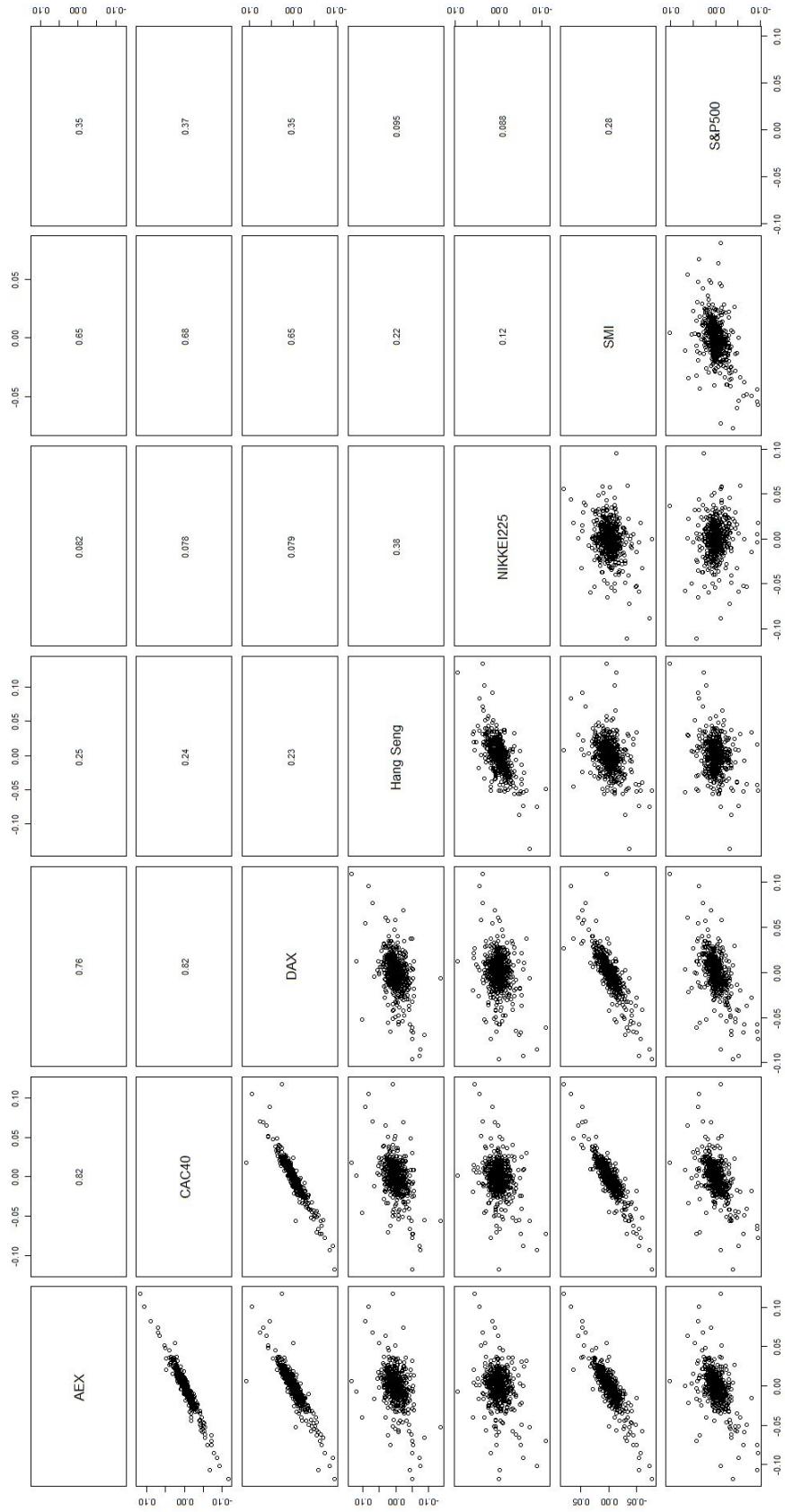


Figure 8: This figure shows scatter plots of the daily log returns between different pairs of some major world indices from the beginning of 2007 until the end of 2009 in the lower panel. In the upper panel the Kendalls  $\tau$  value is reported for all the possible pairs.

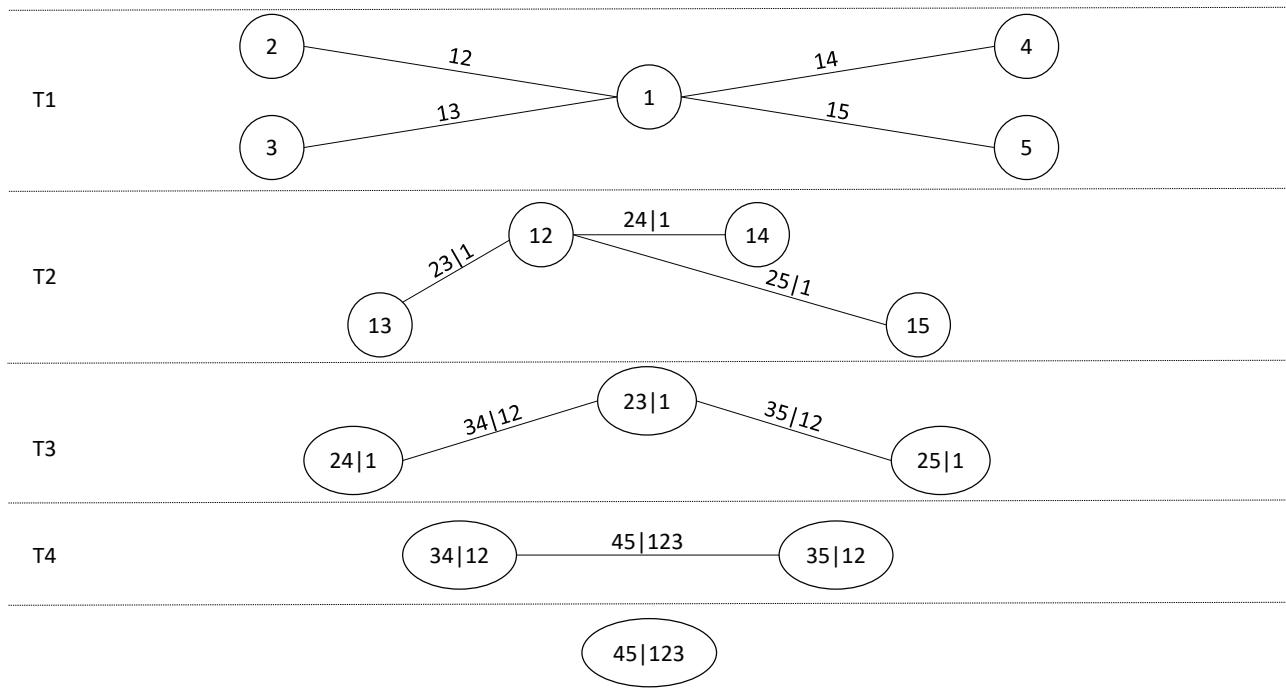


Figure 9: A graphical representation of a C-Vine with five variables, resulting into four trees with ten edges.

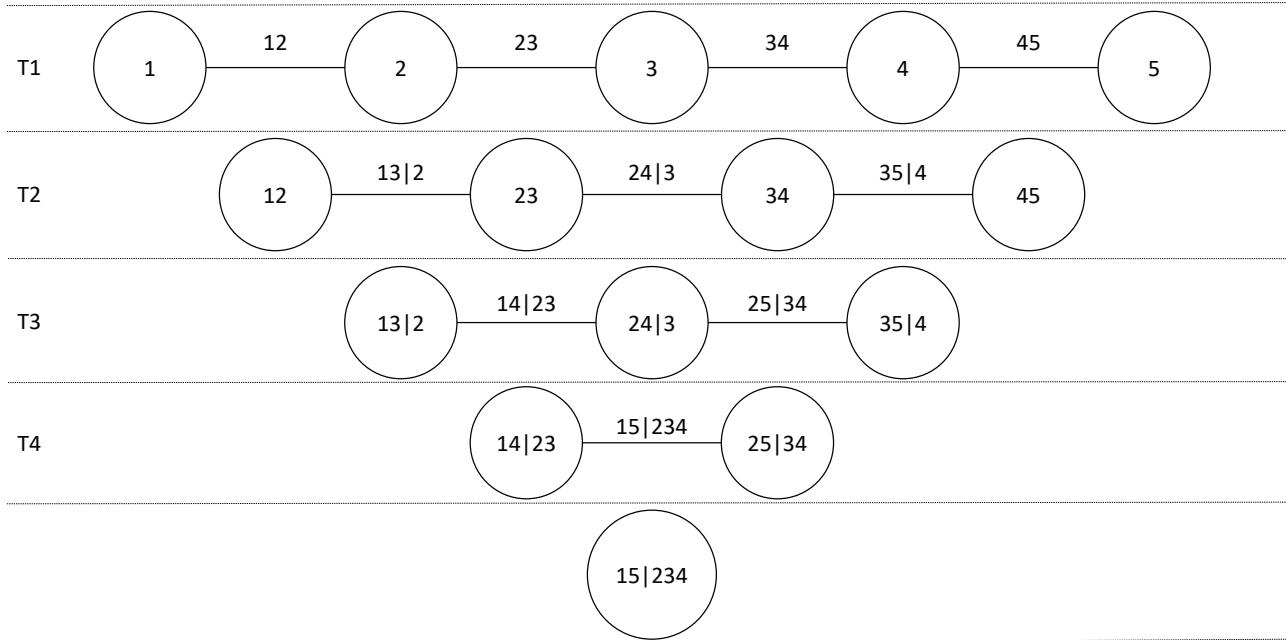


Figure 10: A graphical representation of a D-Vine with five variables, resulting into four trees with ten edges.

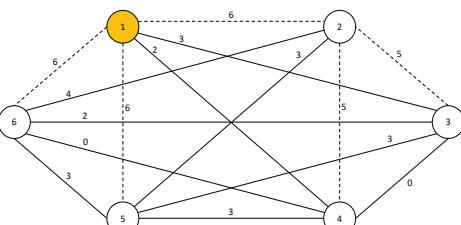
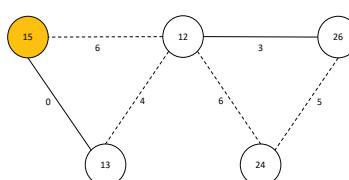
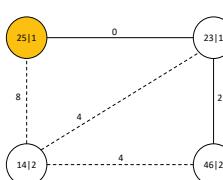
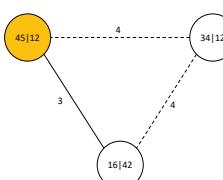
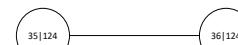
Graph	Description
	This is the starting graph of T1 with six nodes. We choose node 1 to start Prim's Algorithm. The dashed edges are the ones selected by Prim's Algorithm to be in the maximum spanning tree. The values are chosen arbitrarily and should be replaced by the empirical Kendall's $\tau$ .
	Using the maximum spanning tree from T1 this tree (T2) can be created with all possible edges with corresponding Kendall's $\tau$ values. The starting node is 15 and again Prim's Algorithm is used to find the maximum spanning tree.
	The maximum spanning tree from T2 is used to create the tree T3. We use the same strategy as in the previous tree to determine the maximum spanning tree while we choose 25 1 to be the starting node.
	Using the maximum spanning tree from T3 this tree (T4) can be created with all possible edges with corresponding Kendall's $\tau$ values. The starting node is 45 12 and again Prim's Algorithm is used to find the maximum spanning tree.
	This is the last tree that is left. It is not possible to go any further and the R-Vine structure is found using the sequential algorithm.

Table 20: This figure shows an example of using the sequential algorithm on actual trees with arbitrarily chosen values that have to be replaced in practice by values of the empirical Kendall's  $\tau$ .

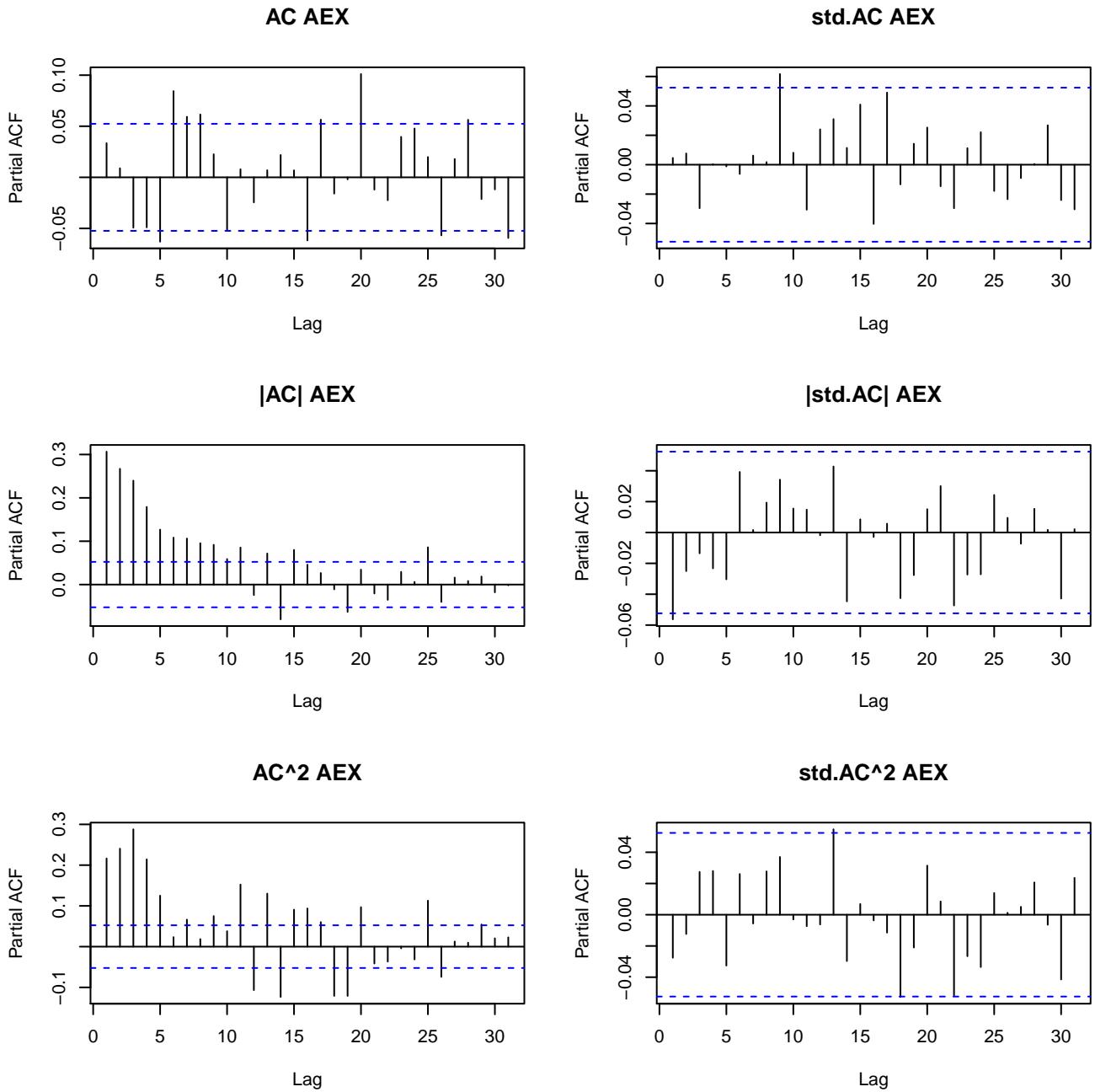


Figure 11: This graph contains several different autocorrelation plots for the AEX dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

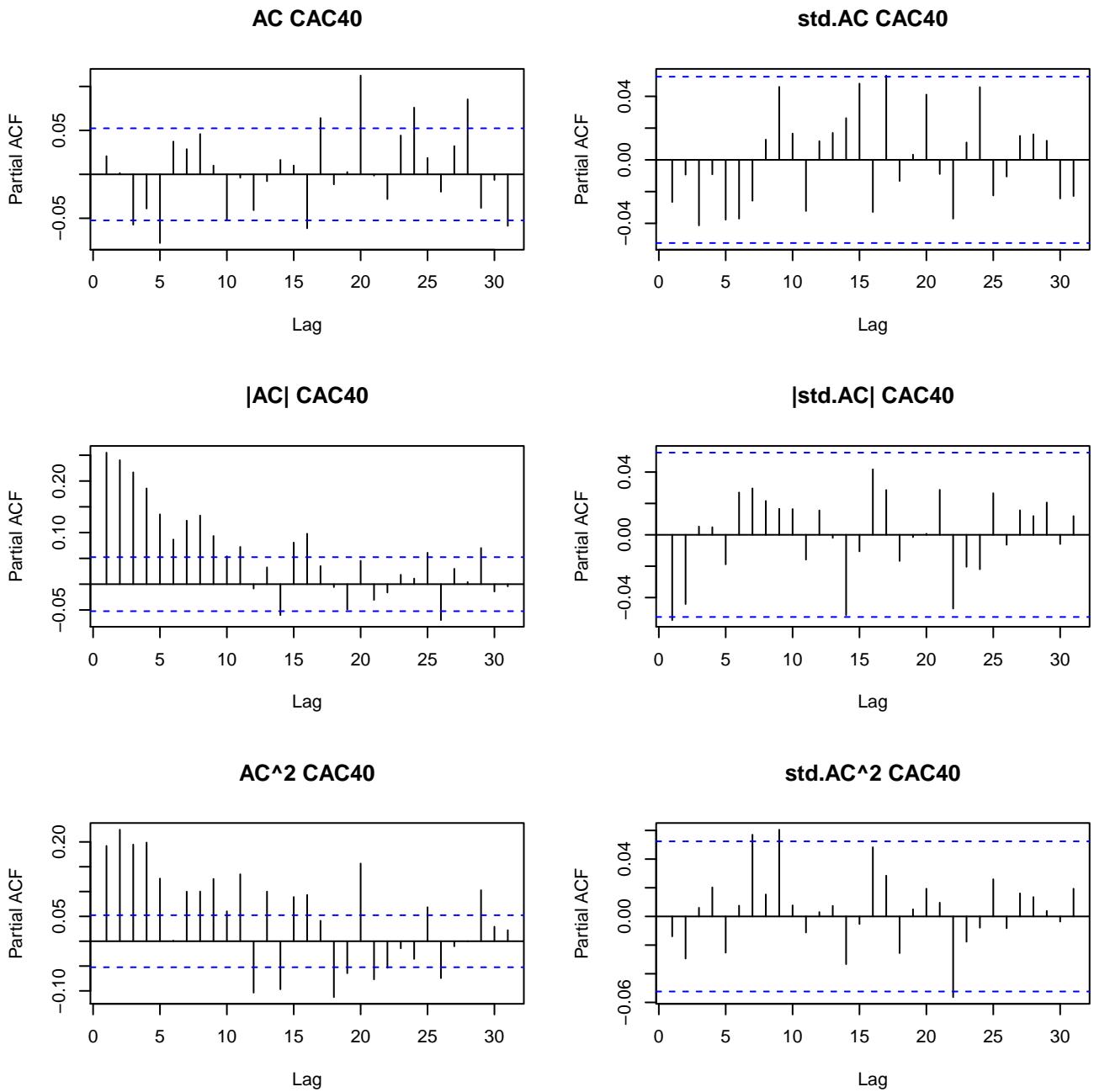


Figure 12: This graph contains several different autocorrelation plots for the CAC40 dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

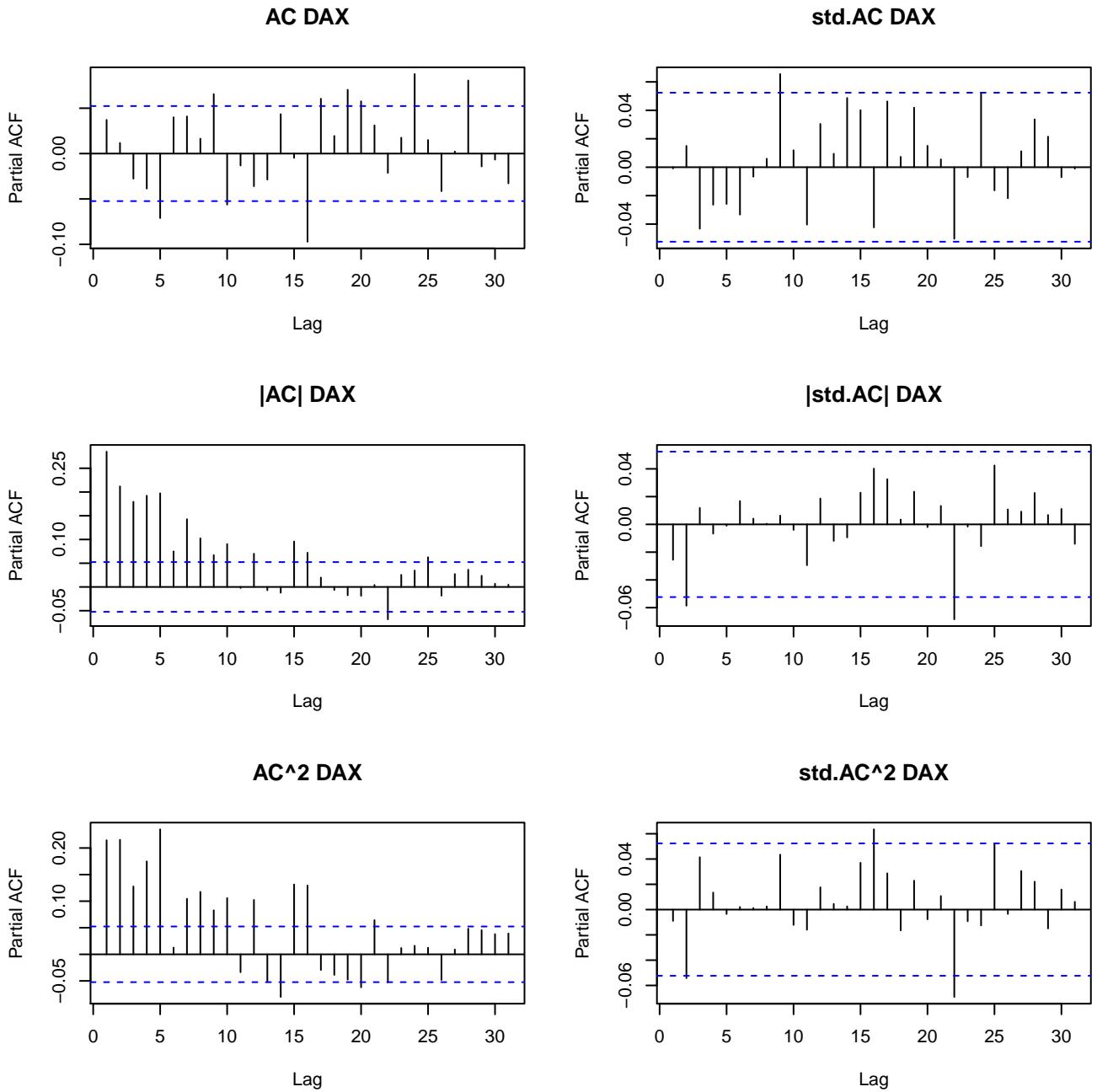


Figure 13: This graph contains several different autocorrelation plots for the DAX dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

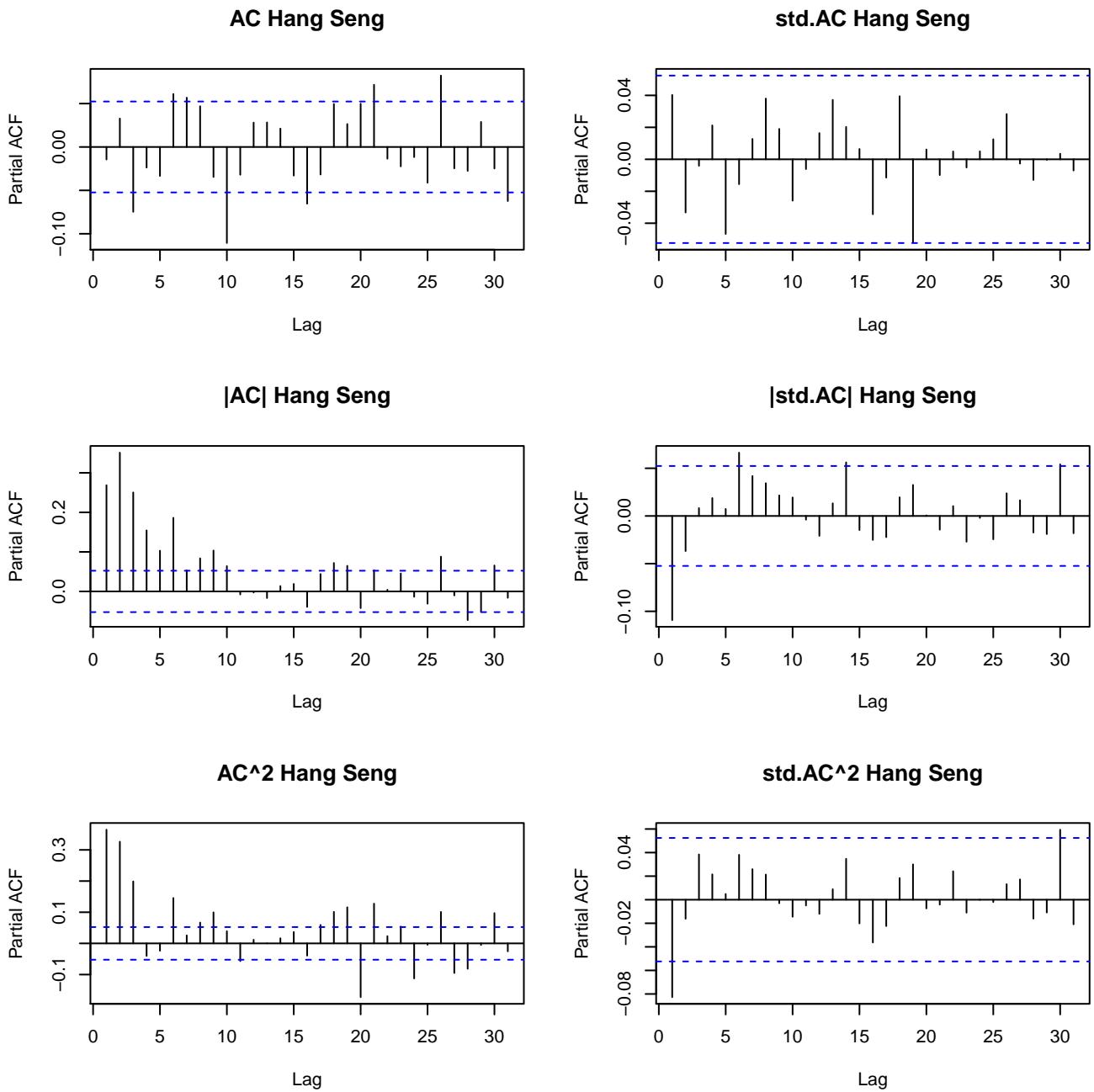


Figure 14: This graph contains several different autocorrelation plots for the Hang Seng dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

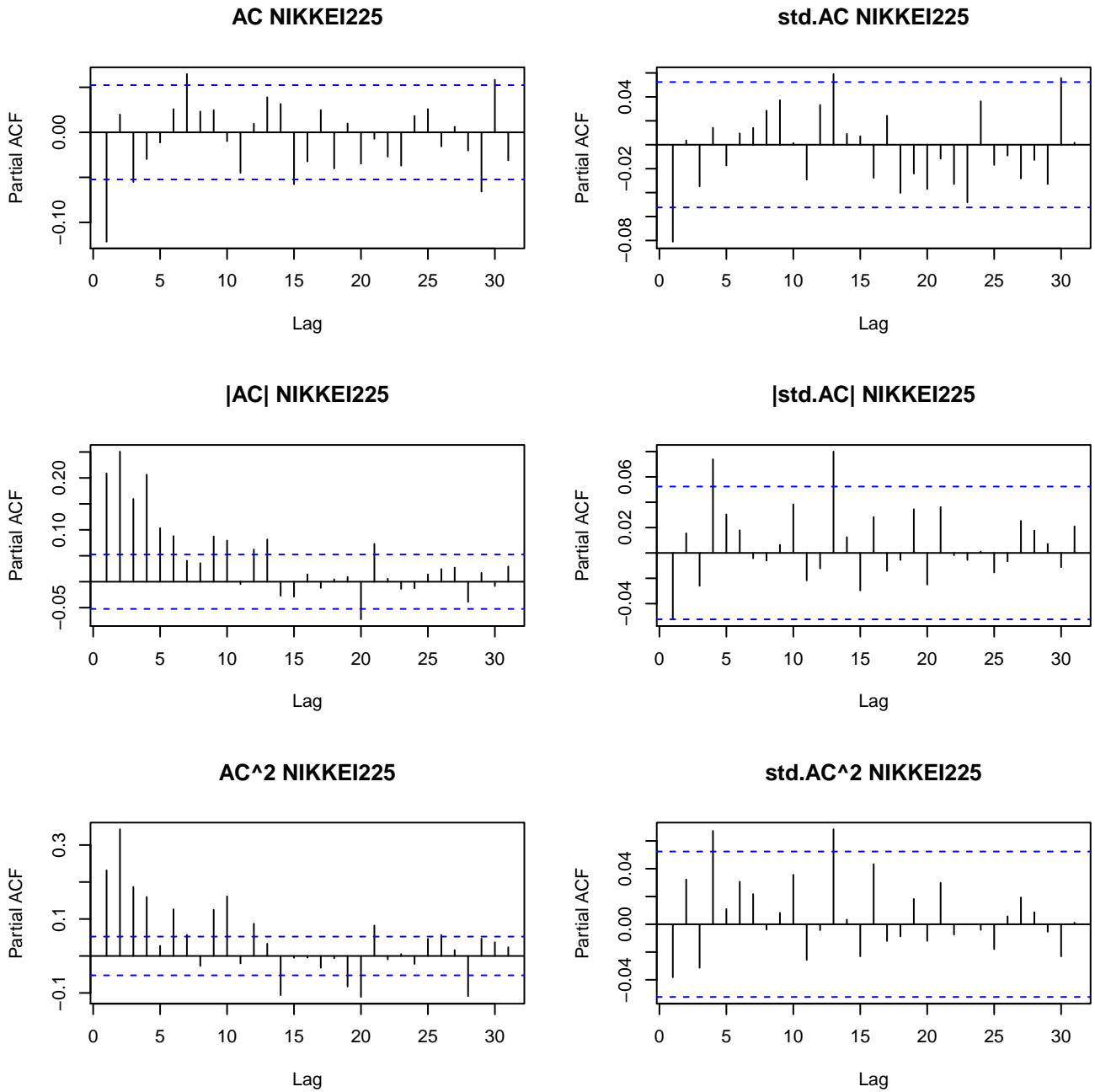


Figure 15: This graph contains several different autocorrelation plots for the AEX dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

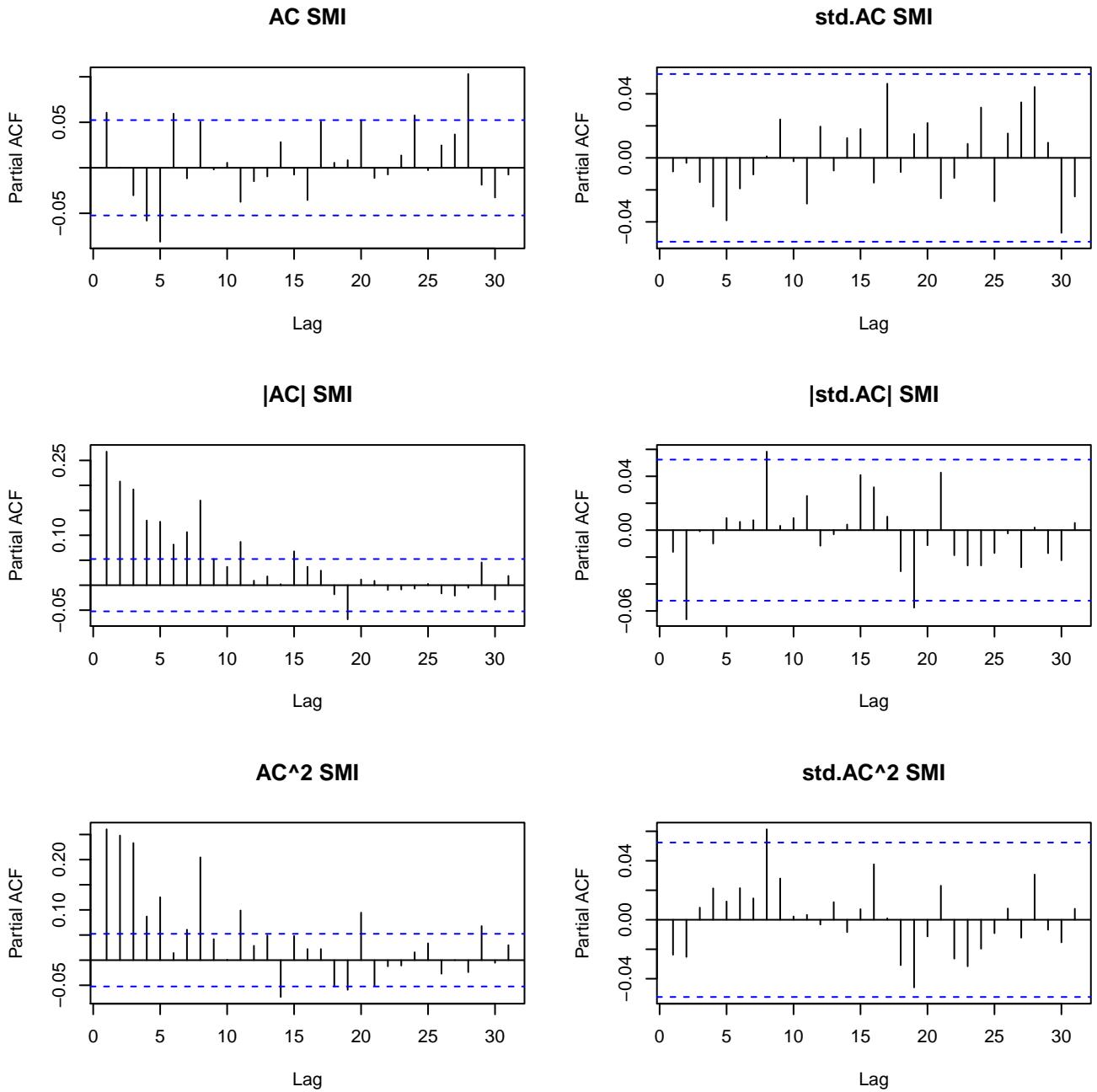


Figure 16: This graph contains several different autocorrelation plots for the SMI dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

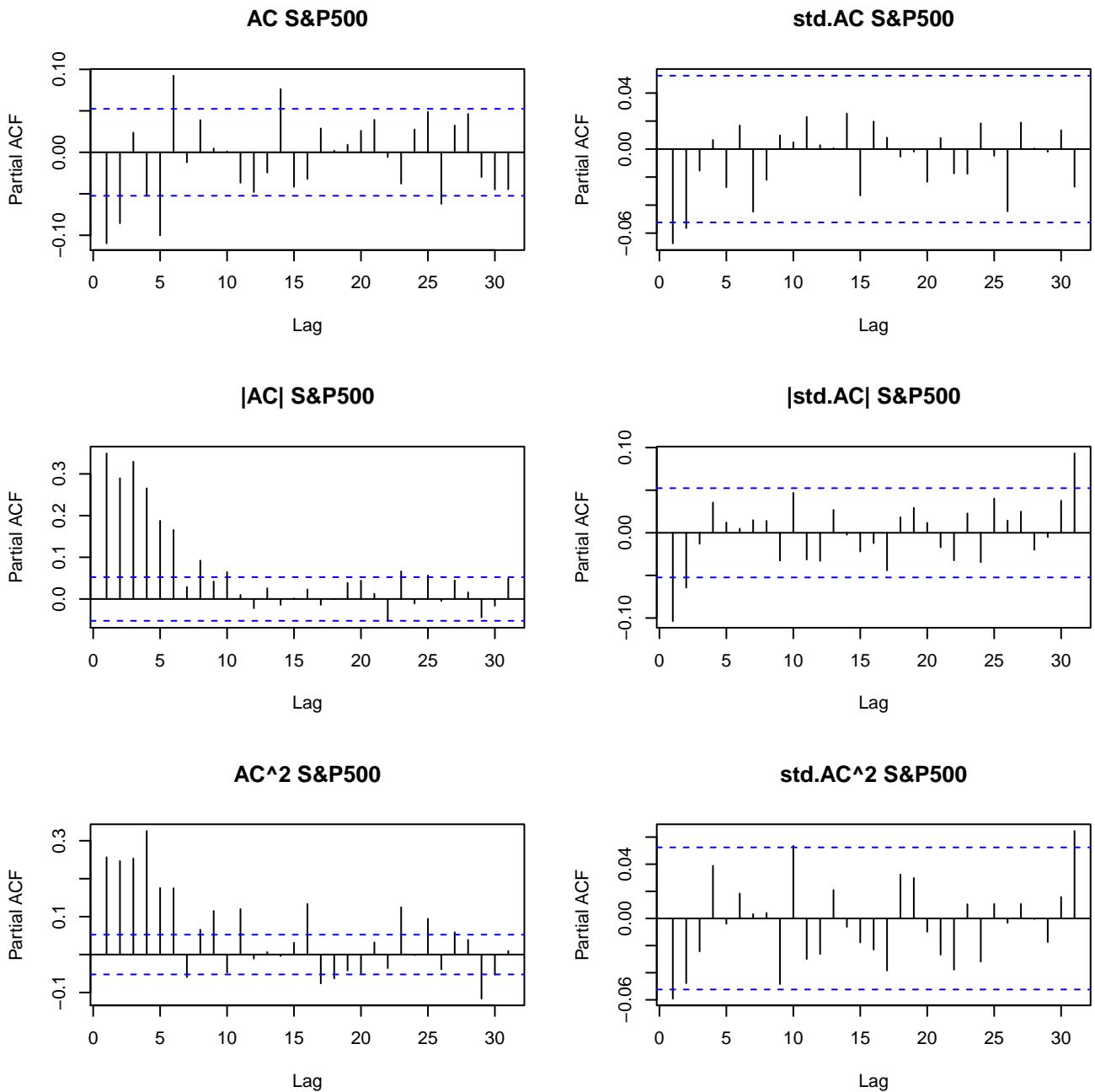


Figure 17: This graph contains several different autocorrelation plots for the S&P500 dataset, containing 1500 daily observations starting in 2002-07-08 until 2009-06-29. The left plots contain the non-filtered autocorrelation coefficients and the right plots contain the filtered ones for the series, the absolute series and the squared series.

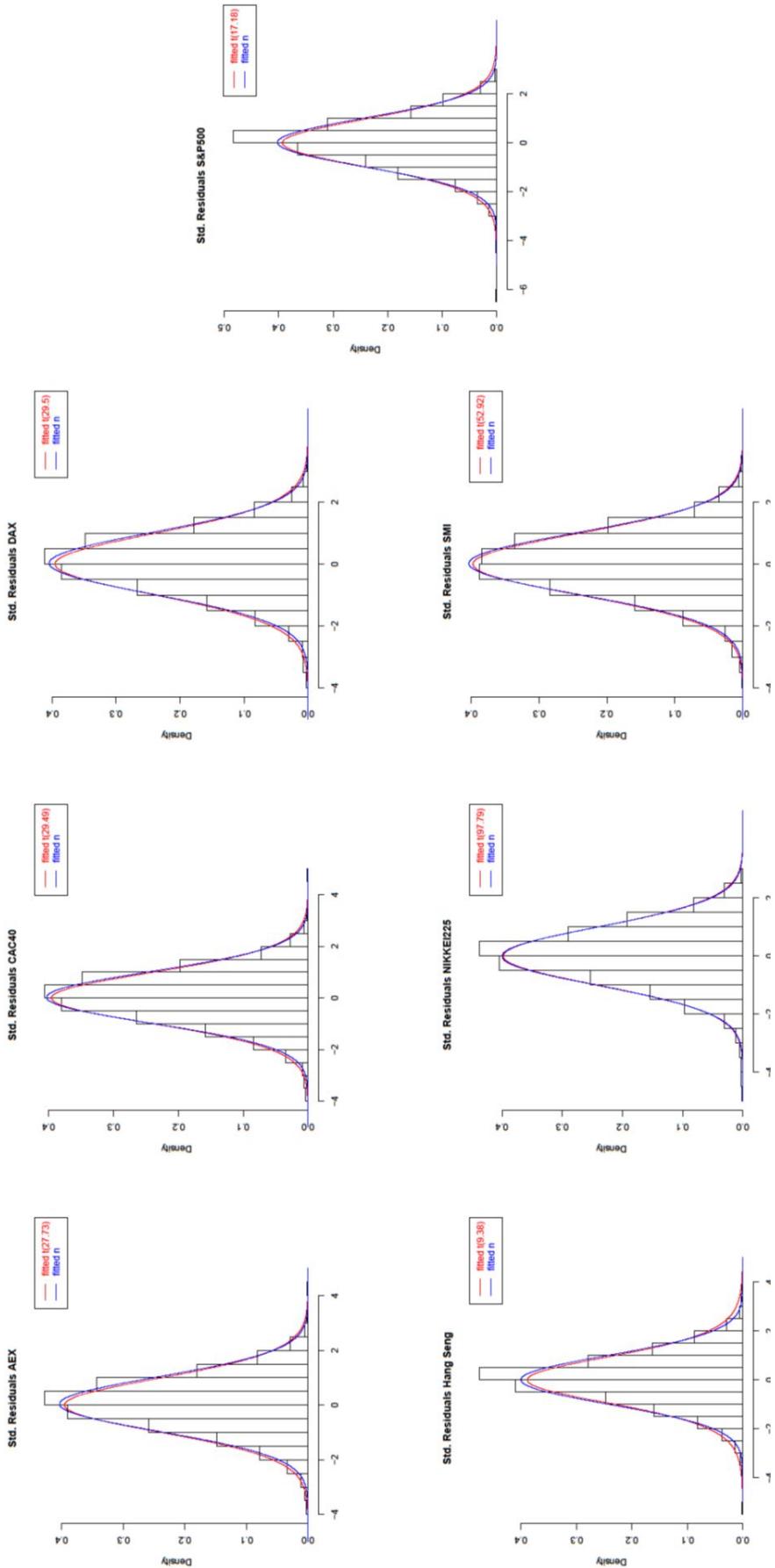


Figure 18: This graph contains several different histograms for all the filtered data of the indices starting in 2002-07-08 until 2009-06-29, containing 1500 observations, along with the fitted student  $t$  with the degrees of freedom within the brackets and the normal distribution.

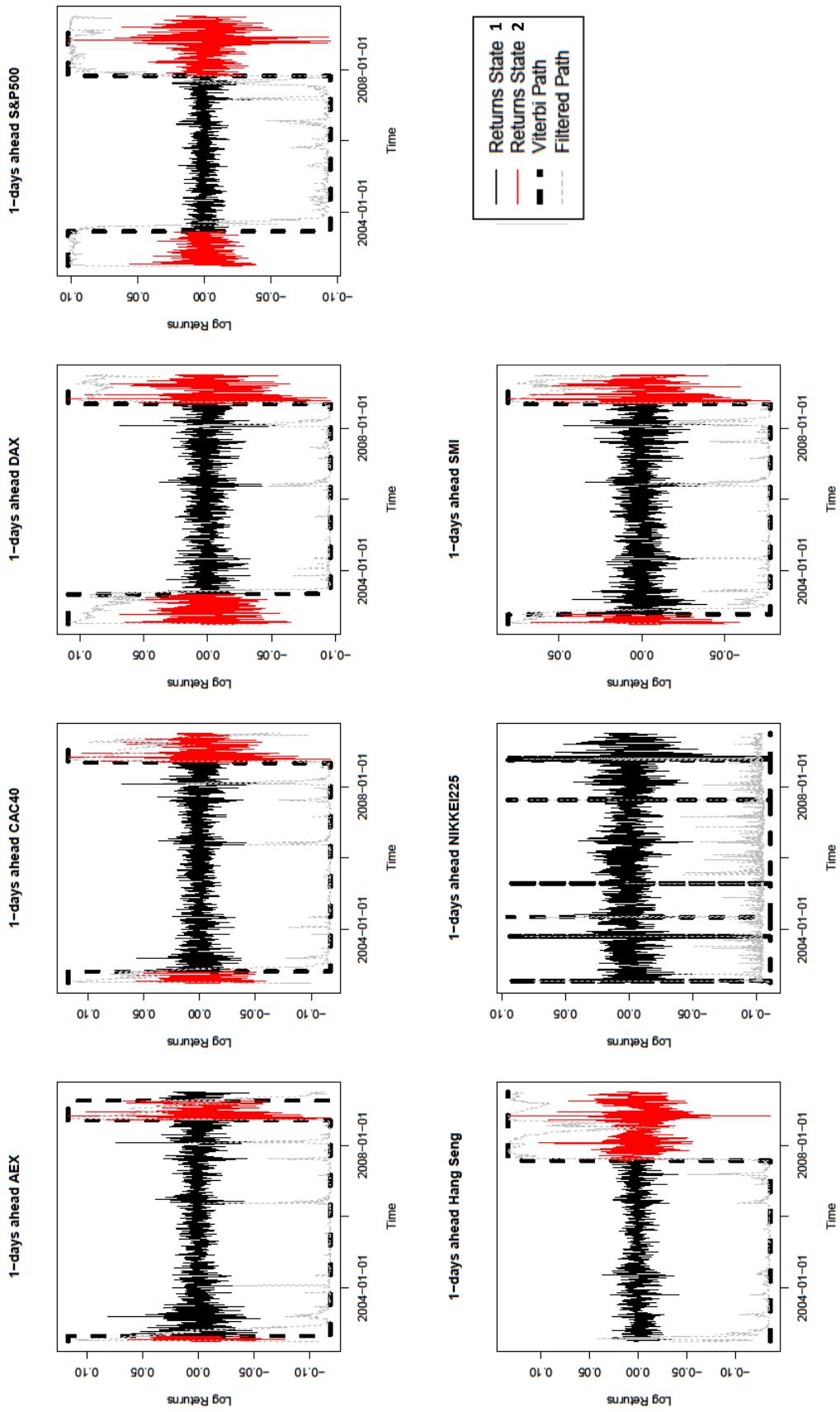


Figure 19: This graph contains several different daily log series the indices dataset starting in 2002-07-08 until 2009-06-29, containing 1500 observations. The black and the red lines correspond with the actual returns. The dashed line represents the Viterbi path. When the Viterbi path is high then the Viterbi path is in state 2 and when it is low in state 1. The gray line represents the filtered path probability scaled to the maximum and the minimum log return.

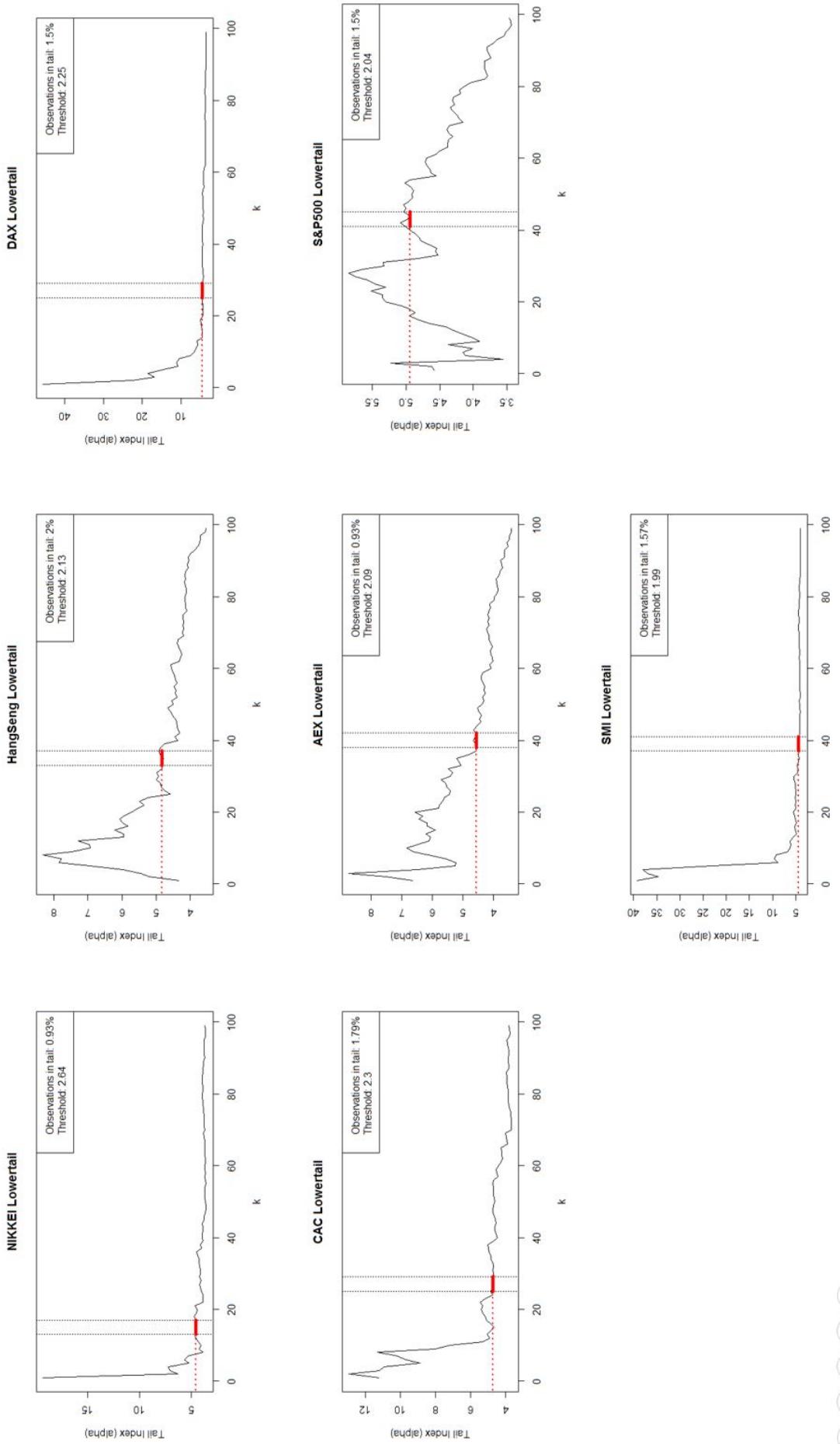


Figure 20: This figure shows the choice of the lower threshold selection algorithm (4.1.2) for the indices dataset using a moving window containing 1500 observations, starting at 18-07-2000 until 01-07-2007. The vertical lines with the red colored segment in between indicate the stable region found by the threshold selection algorithm.



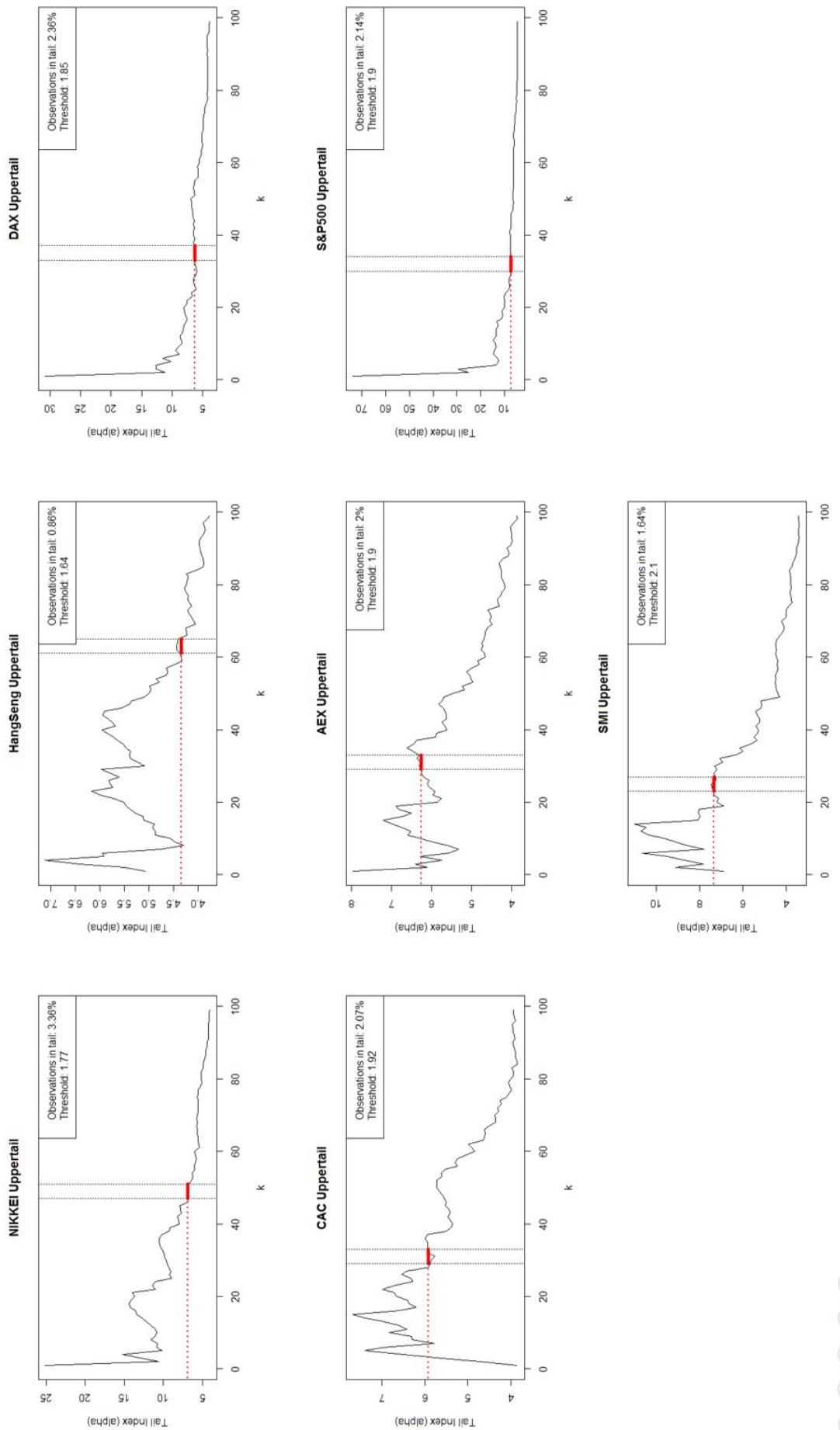


Figure 21: This figure shows the choice of the upper threshold selection algorithm (4.1.2) for the indices dataset using a moving window containing 1500 observations, starting at 18-07-2000 until 01-07-2007. The vertical lines with the red colored segment in between indicate the stable region found by the threshold selection algorithm.



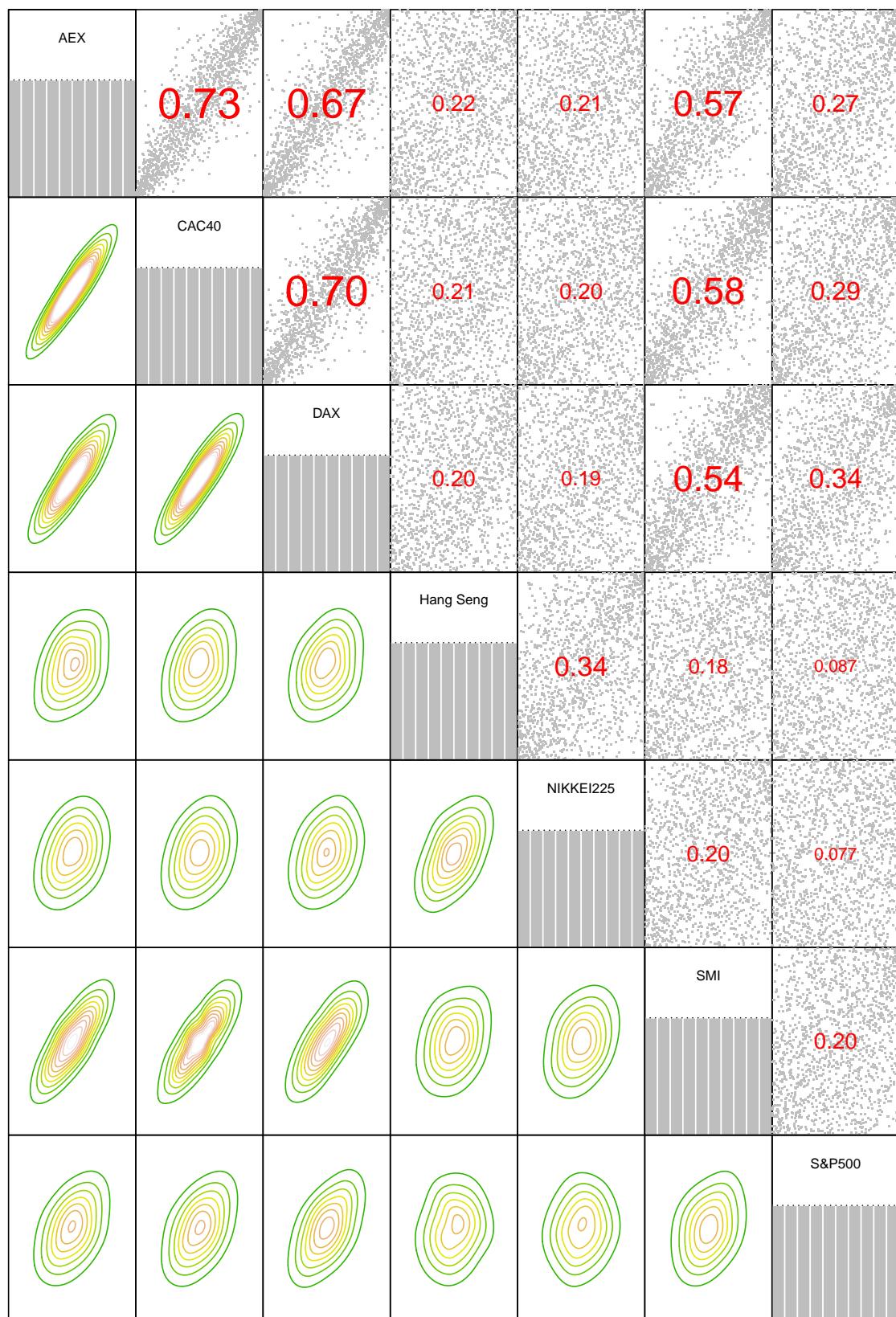


Figure 22: This figure contains the pair scatter plots (upper triangle) and the pair contour plots (lower triangle) of a datasets consisting 1500 observations until 2007-07-01.

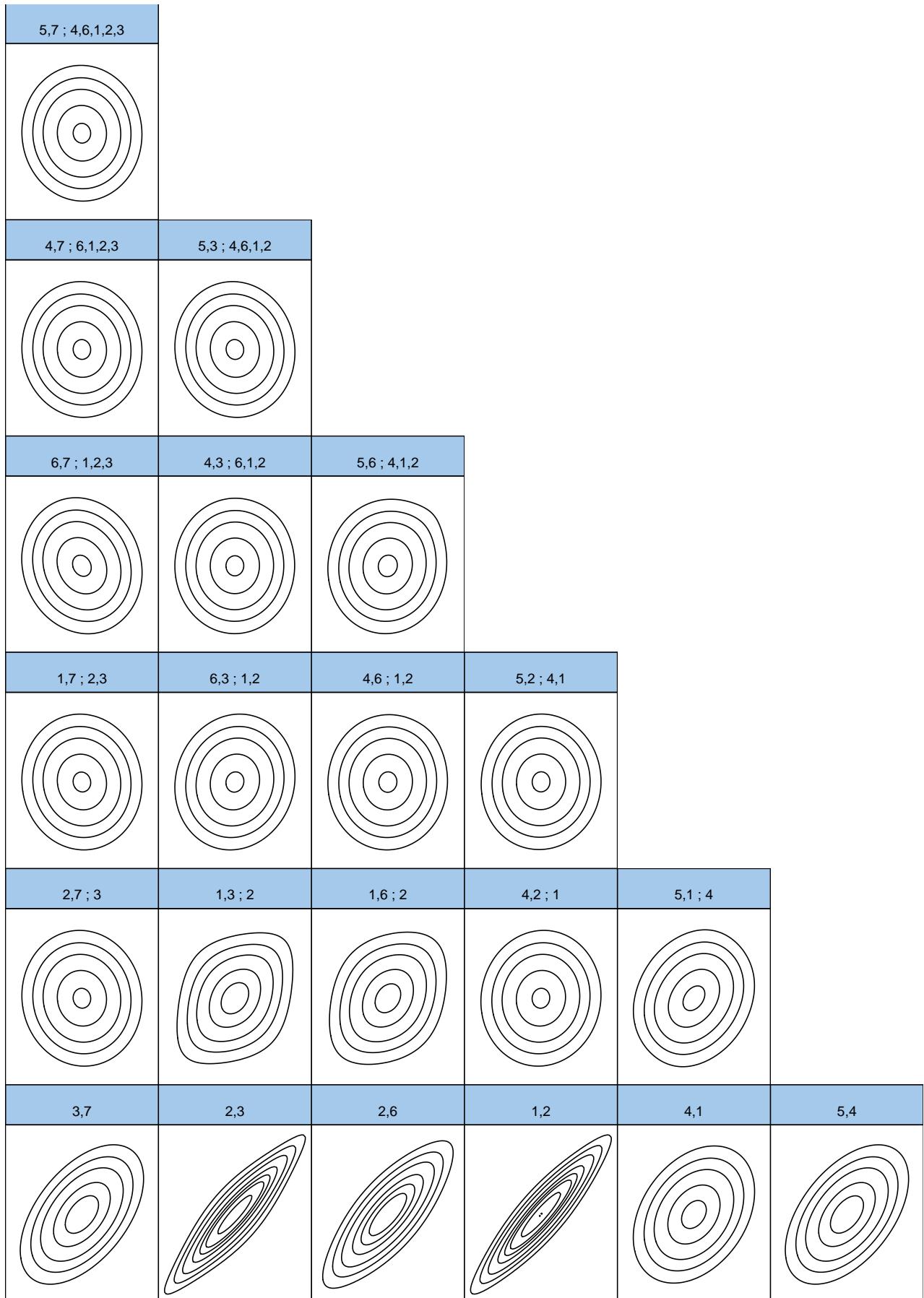


Figure 24: This figure contains by the Vine Copula implied contour plots of 23

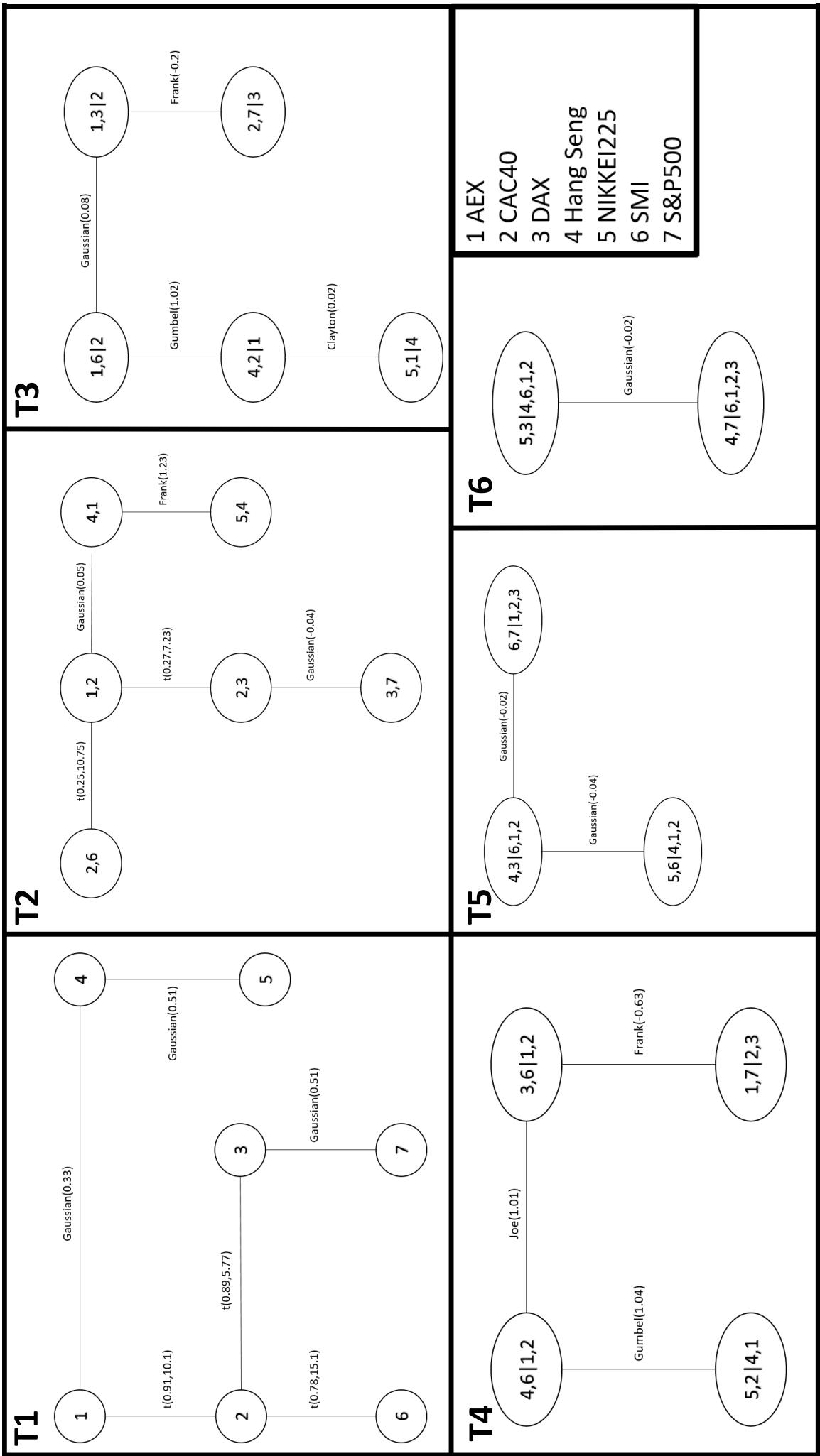


Figure 23: This figure contains the estimated Vine Copula on 2007-07-01 using a window of 1500 observations. Each panel represents a tree and the edges represent the estimated bivariate copula dependence along with the estimated parameter.

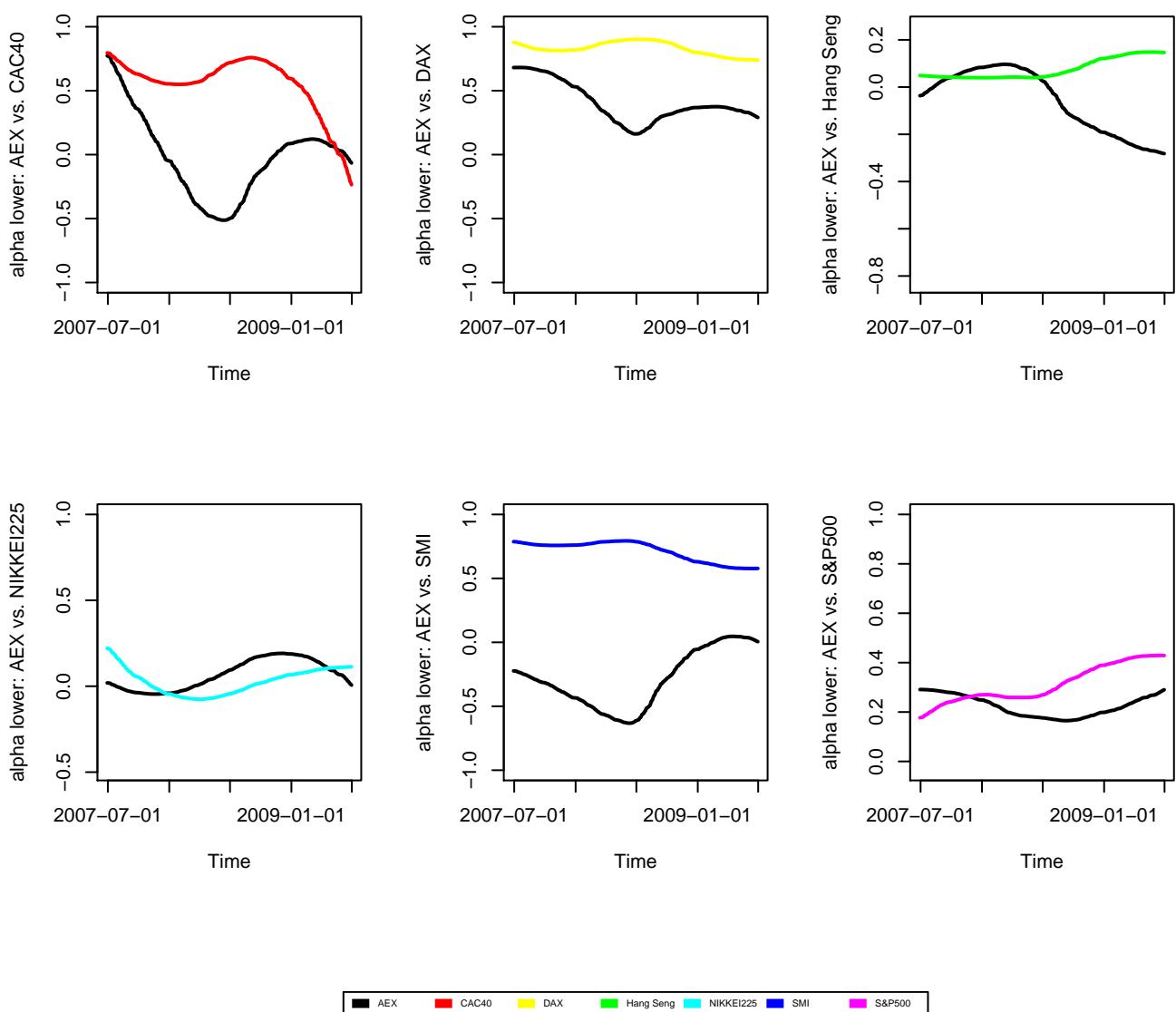


Figure 25: This figure contains the estimated parameter  $a^-$  of the MEVT of all the indices with respect to the AEX from 2007-07-01 until 2009-07-01.

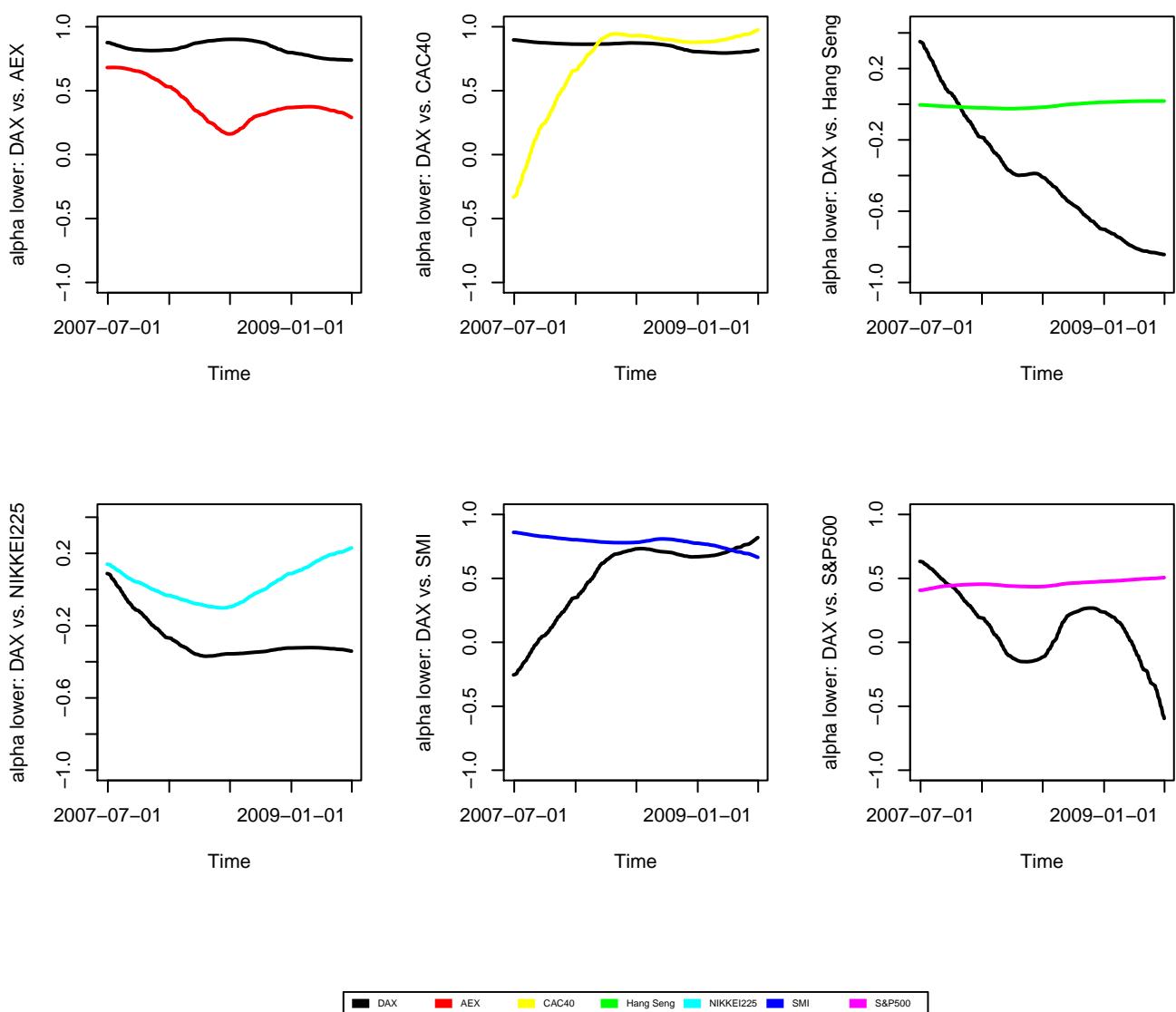


Figure 26: This figure contains the estimated parameter  $\alpha^-$  of the MEVT of all the indices with respect to the DAX from 2007-07-01 until 2009-07-01.

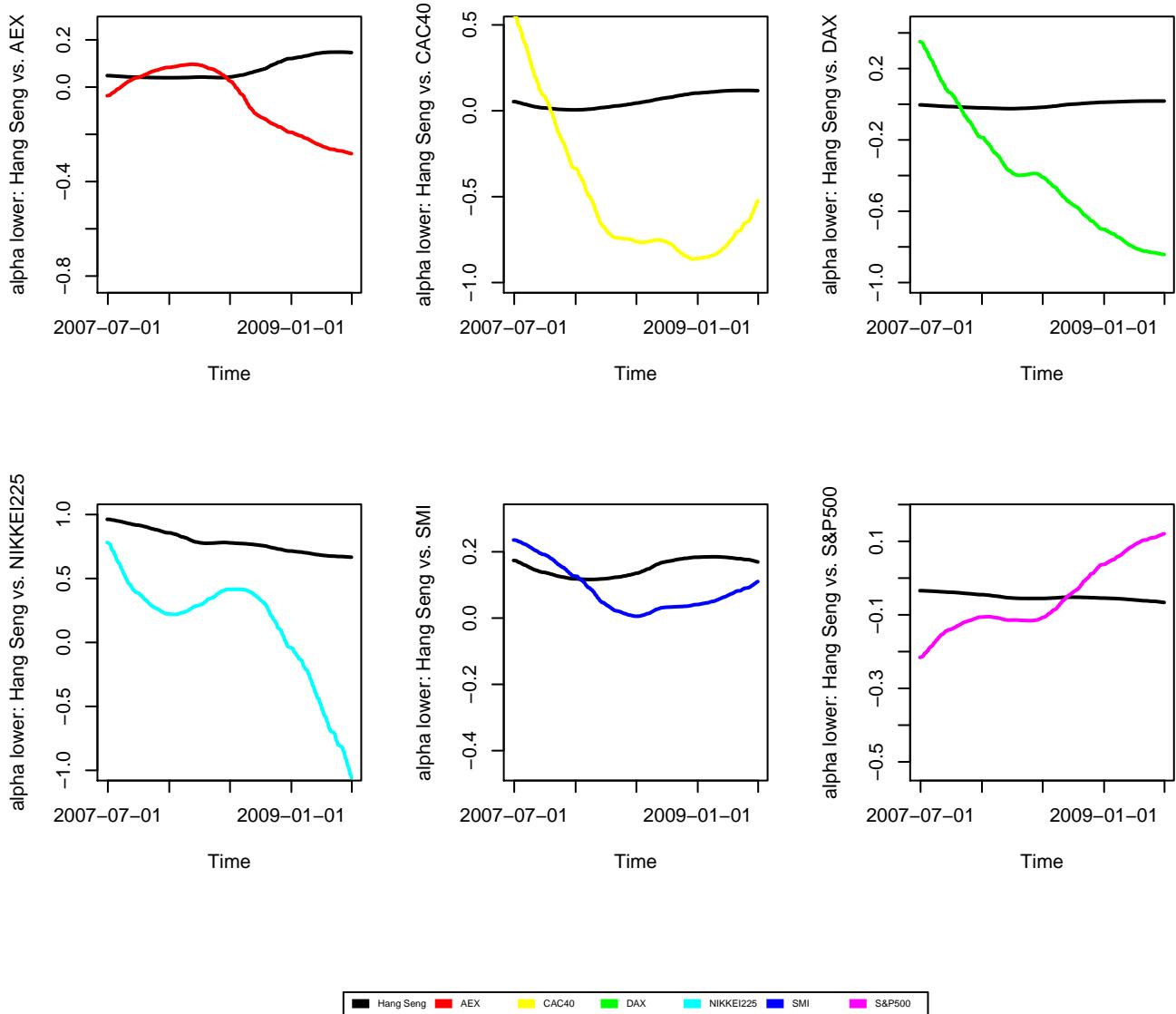


Figure 27: This figure contains the estimated parameter  $a^-$  of the MEVT of all the indices with respect to the Hang Seng index from 2007-07-01 until 2009-07-01.

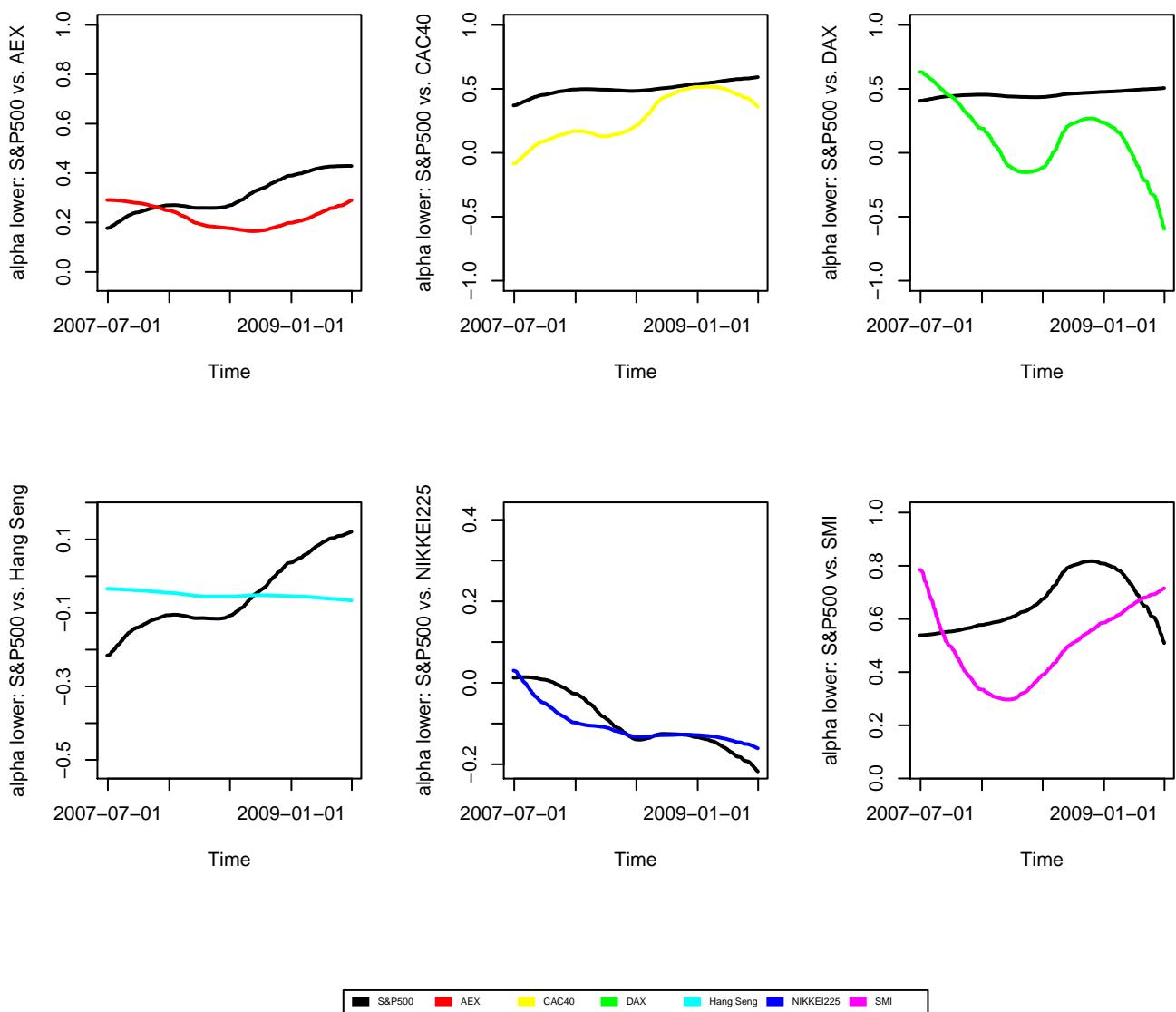


Figure 28: This figure contains the estimated parameter  $a^-$  of the MEVT of all the indices with respect to the S&P500 from 2007-07-01 until 2009-07-01.