École Polytechnique Fédérale de Lausanne

Master's Thesis

the structure of extreme events in the sectors of the S&P 500

Mongi Nouira - Applied Mathematics

July 11, 2022

Abstract

Extreme value statistics provides accurate estimates for the probabilities of rare events such as huge losses in the equity market. We are interested in the extremal dependence between the different sectors of the S&P 500. We perform an exploratory analysis of the S&P 500 sector indices daily losses, using twenty years of data provided by Standard & Poor's. We extract a structure of extreme events using several extremal dependence measures. We employ the framework of regular variation to construct a tail dependence matrix. We present a clustering approach to reveal extremal dependence and detect patterns in extremal observations. We study a conditional model describing the extremal dependence between the S&P 500 and its sectors. One concrete application in mind is to use the extracted dependence structure to suggest a portfolio rebalance, if the investor believes in a future correction and does not consider getting out of the market.

Keywords— Extreme Value Theory, Generalized Pareto Distribution, Angular Distribution, Extremal Dependence, Graphical Lasso, Clustering, Conditional Distribution, Stock Market, S&P 500

Contents

1. Introduction

The year 2021 may be the most surprising one of the stock market in recent history. After the COVID-19 recession and the February 2020 stock market crash, most indices of the United States stock market quickly went back to their all-time high in August 2020, and then entered a rally that kept going higher and higher. In particular, the S&P 500, a stock market index tracking the performance of 500 large companies listed on stock exchanges in the U.S., lost 30% of its value in a couple of months then went on rally where the index doubled its value in 2021. In 2022, a different game is being played, where new rules are set by inflation, supply chain constraints, the raise of interest rates by the U.S. Federal Reserve, overpriced stocks compared to their earnings, high expectations on earnings growth and the war between Russia and Ukraine. Another observation, that may not seem as significant if the past history of the market is ignored, is the very high interest in the stock market from new investors, most of whom are beginners and can trade with only a swipe of a finger on their phone. All these factors have combined to introduce a lot of volatility in the market, which some believe will cause a future recession. Another trend, that is more and more popular among new investors because of the so-called financial experts on social media, but actually makes some sense. This is to simply remove all the hard work of selecting stocks and balancing a portfolio by buying an Exchange-Traded Fund (ETF) that tracks a market index like the S&P 500. Most of these funds are cheap, very liquid and adequately diversified. Another argument usually presented is the power of compound interest: The S&P 500 has an annual compounded rate of return of more than 10% so investing 1000$ each month for 30 years becomes more than 2 million dollars while the initial investment is 360000$, a nice retirement sum if everything goes as planned. The cost of managing such a fund is not significant. In the case of the S&P 500 the two most popular are VOO with an expense ratio of 0.03%, and SPY at 0.09%. Based

on historical data, inflation is rarely a contradiction to this reasoning, since it is not significant compared to the average returns of the market, but it is important to consider.

This report concerns the indices that make up the S&P 500, and represent its different sectors. These can also be traded like any other ETF and can reduce the risk of buying specific companies, but are exposed to the risk associated with the sector. In particular, we want to answer the following questions: What is the dependence structure of these sectors? How does the market depend on its sectors? How do they behave during corrections or large losses? Is it possible to balance a portfolio of these sectors while beating the market? How should one balance the same portfolio during periods of trouble? What can we learn from the historical prices of these sectors? Investing in sectors instead of the whole market may be more appropriate for investors with higher risk appetites, but who do not want to build a portfolio of stocks from scratch. The portfolio of sectors can then be balanced depending on macro-economic news. Another possibility is to extend the ideas behind passive investments, as described by Benjamin Graham and Warren Buffet, who suggest balancing a portfolio between bonds and the market index. With a little more effort, one can balance between bonds and sector indices.

Since the main focus is rare periods of corrections, high volatility and large movements in the market, extreme value theory is the appropriate mathematical framework for this analysis. It is a statistical discipline that describes the unusual rather than the usual. It is commonly used in risk analysis to compute value-at-risk, but there are few applications in the multivariate case, which is more complicated than the univariate or bivariate cases since we have to decide what is considered to be an extreme observation and understand the tail dependence structure. Moreover, computation and validation difficulties can arise due to the curse of dimensionality. The theory of multivariate extremes is well-developed and the literature is still growing. There is a strong interest from mathematicians in this field and many papers of high quality are published each year. We hope to explore many of them to apply their techniques to the subject chosen for this project, and answer most of the previous questions. The next chapter provides a short introduction to extreme value theory, and presents some statistical techniques explored in the report.

2. Background

In this chapter we provide a short introduction to extreme value theory. We start with the univariate case, then turn to the multivariate case, which is the main focus of this thesis. The order of the definitions follows what appears in later chapters. We present the background in a separate chapter in order to have a self-contained report and for clarity. The following is mostly composed of material from Coles (2013), Engelke and Ivanovs (2021) and Hornik et al. (2012).

Extreme value theory is a branch of statistics that seeks to estimate the probability of events that are more extreme than any previously observed. Two main approaches exist for practical uses, namely component-wise block maxima and threshold exceedances. It is widely used for risk assessment on financial markets.

2.1 Univariate EVT

Suppose we have an i.i.d sample $X_1, \dots, X_n$ copies of a random variable X whose distribution function F satisfies certain continuity properties. Let $M_n = \max\{X_1, \dots, X_n\}$. Allowing a linear normalization of M_n, we are interested in the limit of

$$M_n^* = \frac{M_n - a_n}{b_n},$$

for sequences $\{a_n\}$ (re-location), and $\{b_n > 0\}$ (re-scaling), as $n \to \infty$. If there are sequences of constants such that for all $z \in \mathbb{R}$

$$\mathbb{P}(M_n^* \leq z) \to G(z), \quad n \to \infty,$$

where G is a non-degenerate distribution function, then G belongs to the class of generalized extreme value distributions.

Definition 1 (Generalized Extreme Value Distribution)**.** *The generalized extreme value (GEV) distribution is a family of continuous probability distributions that are parameterized by a shape α, location a and scale b parameters. It combines three important families of extreme value distributions known as*

- *Gumbel:* $G(z) = \exp\{-\exp(-\frac{z-a}{b})\}, \quad -\infty < z < \infty,$
- *Fréchet:* $G(z) = \begin{cases} 0, & z \leq a, \\ \exp\{-(\frac{z-a}{b})^{-\alpha}\}, & z > a, \end{cases}$
- *Weibull:* $G(z) = \begin{cases} \exp\{(\frac{z-a}{b})^{\alpha}\}, & z < a, \\ 1, & z \geq a. \end{cases}$

The three families can be combined into a single distribution of the form

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\},$$

where the parameters satisfy $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$, known as the GEV family distribution.

Suppose that $\mathbb{P}(M_n \leq z) = G(z)$, for some $\mu, \sigma > 0$ and ξ. G is a GEV distribution function. Then, for a large enough u, the distribution function of $X - u$ conditional on $X > u$ is

$$H(z) \coloneqq \mathbb{P}(X - u \leq z \mid X > u) = 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)_+^{-1/\xi},$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$ and $z \in \{z : z > 0, 1 + \xi z/\hat{\sigma} > 0\}$. The family of distributions defined by H is called the generalized Pareto family. This means if block-maxima has a distribution function G, then threshold excesses have a corresponding distribution function within the generalized Pareto family.

Definition 2 (Generalized Pareto Distribution)**.** *The generalized Pareto distribution (GPD) is a family of continuous probability distributions, often used to model the tails of another distribution. It has the survivor function*

$$\left(1 + \frac{\xi z}{\tilde{\sigma}}\right)_+^{-1/\xi},$$

where $-\infty < \mu < \infty$ is the location, $\tilde{\sigma} > 0$ the scale and $-\infty < \xi < \infty$ the shape parameters.

Definition 3 (Max-Stable Distribution)**.** *A random variable X is max-stable if for every $n \geq 2$ and i.i.d sample $X_1, \ldots, X_n$ copies of X, there exist constants $c_n > 0$ and d_n such that*

$$M_n = \max\{X_1, \ldots, X_n\} \stackrel{d}{=} c_n X + d_n.$$

The class of max-stable distributions coincides with the class of all possible non-degenerate distributions for normalized maxima of i.i.d. r.v.'s.

Definition 4 (Regular Variation)**.** *A random variable X with distribution F is regularly varying with variation tail index α if $F : \mathbb{R}^+ \to \mathbb{R}^+$ and*

$$\lim_{t\to\infty} \frac{F(tx)}{F(t)} = x^{-\alpha}, \quad x > 0.$$

We write $X \in RV_+(\alpha)$.

The univariate case may not be the priority of the thesis, since we want to analyze multivariate dependence, but it is useful for marginal models and exploratory data analysis. Now we focus on multivariate extreme value theory.

2.2 Multivariate EVT

Suppose we have an independent sample $X^1, \ldots, X^n$ from a random vector $X \in \mathbb{R}^d$. Like in the univariate case, suppose there exist sequences of constants $a^n > 0$ and b^n in $\mathbb{R}^d$ and a distribution function G, such that

$$\mathbb{P}\left(\frac{\max\limits_{i=1,\ldots,n} X_1^i - b_1^n}{a_1^n} \leq x_1, \ldots, \frac{\max\limits_{i=1,\ldots,n} X_d^i - b_d^n}{a_d^n} \leq x_d\right) \to G_X(x), \quad n \to \infty,$$

for all continuity points $x \in \mathbb{R}^d$ of G_X. All the marginal distributions of G_X are univariate extreme value distributions. Let F_i be the distribution function of X_i for $i = 1, \ldots, d$. If we apply the transformation

$$X \to Y = \left(\frac{1}{1 - F_1(X_1)}, \ldots, \frac{1}{1 - F_d(X_d)}\right)$$

to X, then for $y \in [0,\infty)^d$ we have

$$\mathbb{P}\left(\frac{\max\limits_{i=1,\dots,n} Y_1^i}{n} \leq y_1, \dots, \frac{\max\limits_{i=1,\dots,n} Y_d^i}{n} \leq y_d\right) \to G(y), \quad n \to \infty,$$

and

$$G(y) = \exp[-\nu(\{u \in [0,\infty)^d \ : \ \exists i \ u_i > y_i\})],$$

where ν is called the exponent measure.

Definition 5 (Exponent Measure)**.** *The exponent measure ν contains all information on the extremal dependence of Y.*

The exponent measure is homogeneous of order -1, and there exists a constant $c > 0$ such that

$$\nu(\{u \in [0,\infty)^d \ : u/||u|| \in B, ||u|| > z\}) = cz^{-1}S(B),$$

where $B \subset \mathbb{S}_+^{d-1} = \{x \in [0,\infty)^d : ||x|| = 1\}$ is a Borel set, $||.||$ a norm and $z > 0$.

Definition 6 (Angular Measure)**.** *The probability measure S is called an angular (or spectral) measure. We have*

$$\mathbb{P}\left(\frac{Y}{||Y||} \in B \;\middle|\; ||Y|| > z\right) = S(B), \quad z > 0, \ B \subset \mathbb{S}_+^{d-1}.$$

Using the sample $X^1, \dots, X^n$, the angular measure can be empirically estimated by the polar transformation to a radial and angular component

$$R_i = ||X^i||, \quad W_i = \frac{X^i}{R_i}.$$

Definition 7 (Regular Variation)**.** *A random vector is multivariate regularly varying if its joint tail decays like a power function. A random vector X taking values in $\mathbb{R}_+^d$ is regularly varying if there exist a sequence $b_n \to \infty$ and a limit measure ν_x for Borel sets in $\mathbb{R}_+^d \setminus \{0\}$ such that*

$$n\mathbb{P}(b_n^{-1}X \in \cdot) \xrightarrow{v} \nu_x(\cdot),$$

as $n \to \infty$ and $\xrightarrow{v}$ denotes vague convergence in the space of non negative Radon measures on $\mathbb{R}_+^d \setminus \{0\}$. We let $X \in RV_+^d(\alpha)$ denote a regularly varying vector X with tail index α. It can be shown that $b_n = L(n)n^{1/\alpha}$, where L is a slowly varying function, i.e. $L(ct)/L(c) \to 1$, as $c \to \infty$ and $t > 0$.

Definition 8 (Multivariate Max-Stability)**.** *A d-variate distribution function G is called max-stable if for every $k \in \mathbb{N}^*$ and $x \in \mathbb{R}^d$ we can find vectors $\beta_k > 0$ and α_k such that*

$$G^k(\beta_k x + \alpha_k) = G(x).$$

Definition 9 (Pickands Dependence Function)**.** *A measure of asymptotic dependence between two variables, denoted A and can be computed as*

$$A(t) = 1 - t + 2\int_0^t \nu([0, w])\mathrm{d}w, \quad 0 \leq t \leq 1.$$

This formula allows the computation of the exponent measure as

$$\nu([0, w]) = \begin{cases} \frac{1+A'(w)}{2}, & 0 \leq w \leq 1, \\ 1, & w = 1, \end{cases}$$

where A' is the derivative of A.

The transformation most used in the following chapters is the re-scaling of the margins to standard Fréchet(α), i.e. $G(x) = \exp(-x^{-\alpha})$ for $x > 0$ and $\alpha > 0$. In that case, we know that G is max-stable and its margins are univariate max-stable. This transformation is useful as a pre-processing step to estimate dependence measures and fit some extreme value models.

2.3 Statistical Techniques

Principal Component Analysis

Principal Component Analysis (PCA) is an orthogonal linear transformation that transforms the data to a new coordinate system, allowing dimension reduction. Let $X \in \mathbb{R}^{n\times d}$ be a data matrix, $\Sigma \in \mathbb{R}^{d\times d}$ its covariance matrix, and $\mu \in \mathbb{R}^d$ its mean. We denote $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, the sorted eigen-values of Σ, and $u_1, \ldots, u_p$ the corresponding eigen-vectors, called the loadings, and play the role of a map between the original observations and the new coordinate system. The i-th principal component observation, called the i-th score, is defined as $Y_i = (X - \mathbf{1}_n\mu')u_i$. The scores are the new representation of the original data, and allows to reduce its dimension by selecting the first p scores, where $1 \leq p < d$ (Jolliffe 2002).

Partial Correlation

Partial correlation measures the degree of association between two random variables after controlling for linear effects of known confounding variables in order to avoid misleading results, which are common for multivariate data when using the correlation coefficient. The partial correlation between X and Y given a set of controlling variables $Z = \{Z_1, \ldots, Z_p\}$ is the correlation between the residuals resulting from fitting a linear model of X with Z and Y with Z, denoted e_X and e_Y respectively. The partial correlation is

$$\rho_{XY \cdot Z} = \frac{\langle e_X, e_Y \rangle}{||e_X||_2 ||e_Y||_2}.$$

Another way to compute the partial correlations is using matrix inversion. Let $V = \{X_1, \ldots, X_d\}$ a set of random variables, Σ their covariance matrix and $K = \Sigma^{-1}$ the precision matrix. We have

$$\rho_{X_i X_j \cdot \mathbf{V} \setminus \{X_i, X_j\}} = -\frac{K_{ij}}{\sqrt{K_{ii} K_{jj}}}, \quad 1 \leq i \neq j \leq d.$$

Partial correlation is used as a measure of conditional independence (Wang 2013).

Gaussian Graphical Model

A graph is a structure composed of a set of objects called vertices or nodes. A pair of nodes can be connected by an edge. Weights can also be associated with each edge. We write $G = (V, E)$ or $G = (V, E, W)$, where V is the set of vertices, E the set of edges and W the set of weights. A graphical model uses graphs as a representation of the conditional dependence structure between random variables. A Gaussian graphical model is a special type of such graphical model. Let $X \sim \mathcal{N}(\mu, \Sigma)$ be a random vector in $\mathbb{R}^d$ following a multivariate normal distribution. Let $K = \Sigma^{-1}$, called the precision matrix. The structure of the graphical model is $G = (V, E)$, where $V = \{1, \ldots, d\}$ and $(i, j) \in E \iff K_{ij} \neq 0$. The precision matrix can be estimated using maximum likelihood, but the estimator is not sparse with probability 1. Sparsity can be introduced in a graphical model by using the penalized maximum likelihood solved by the graphical lasso, in order to estimate the precision matrix K. The penalized maximum likelihood estimator can be solved using the graphical lasso algorithm (Banerjee, El Ghaoui, and d'Aspremont 2008; Friedman, Hastie, and Tibshirani 2008).

k-means Clustering

k-means clustering is used to partition observations into k clusters. It outputs centroids that represent the cluster centers, and labels that assign the most probable cluster to each observation. It is generally based on the Euclidean distance and fitted recursively. Let $x_1, \ldots, x_n$ observations in $\mathbb{R}^d$. The objective is to find $S = \{S_1, \ldots, S_k\}$ solution of

$$\arg\min_{S} \sum_{i=1}^{k} \frac{1}{|S_i|} \sum_{\mathbf{x},\mathbf{y} \in S_i} \|\mathbf{x} - \mathbf{y}\|^2 .$$

There exist other variations of k-means, for example a different distance measure (Hornik et al. 2012).

The next chapter provides an exploratory data analysis, where the priority is extreme observations, i.e. large losses in the S&P 500 sectors. We present the dataset, we select an appropriate threshold to classify losses as extreme, then we analyze the pairwise dependence between sectors based on several measures.

3. Exploratory Analysis

3.1 Data

We are interested in the different sectors of the U.S. stock market as defined by the Global Industry Classification Standard (GICS). We use the indices provided by Standard & Poor's for the S&P 500 market. There are 11 sectors and one global index:

1. Market (S&P 500)
2. Communication Services
3. Technology
4. Industrial
5. Materials
6. Consumer Discretionary
7. Financial
8. Health Care
9. Consumer Staples
10. Utilities
11. Real Estate
12. Energy

We included the whole S&P 500 market index to understand the tail dependence structure of losses between the market and its sectors. Since some indices are newer than the others, we had to remove some observations to have a complete dataset with all sector indices available each day. We deal with a small number of missing values using forward propagation. In the end we have 5064 observations for 12 indices corresponding to the period from 11/02/2002 until 22/03/2022. This should be sufficient for the analysis of extremes since it includes major crises such as the oil price bubble, the 2008 financial crisis, the COVID-19 crisis and the start of the 2022 Russia/Ukraine war.

For the remaining part of the report, we consider the scaled log daily losses defined in terms of the adjusted close prices of an index P_t as

$$ll_t = -100 \log\left(\frac{P_t}{P_{t-1}}\right).$$

3.2 Threshold Selection

In this section, we define what constitutes an extreme event, i.e. how large should a daily loss be in order to be considered as extreme. We start with a univariate analysis of the variables corresponding to sectors. In particular, we are interested in threshold selection for extreme value analysis. Two main methods are an exploratory technique carried out prior to model estimation and an assessment of the stability of parameter estimates based on fitting models across a range of different thresholds (Coles 2013).

The first method involves fitting a GPD with parameters σ and ξ to the excesses at different thresholds u. If the GPD is valid as a model then the means of the excesses are expected to change linearly with u. This leads to the **mean residual life plot** defined as

$$\left\{\left(u, \frac{1}{n_u}\sum_{i=1}^{n_u}(x_i - u)\right) : u < x_{\max}\right\},$$

where $x_1, \ldots, x_{n_u}$ are the observations that exceed u and $x_{\max}$ is the largest of the x_i. Confidence intervals can be computed based on approximate normality of sample averages.

Figure 3.1 contains the mean residual life plots for each sector. The plots are linear for negative values, and then curvature appears beginning from -1. They are reasonably linear after $u = 2$ but differ in shape depending on the sector. One can argue that the threshold $u = 2$ is too low for sectors such as Financial or Real Estate or too high for sectors such as Consumer Staples and Health Care (no confidence intervals are estimated for high thresholds because n_u is too small). This will be investigated later to make sure results, interpretations and dependence metrics do not drastically change when increasing the threshold. The threshold $u = 2$ leads to between 100 and 400 exceedances depending on the sector as seen in Figure 3.2.

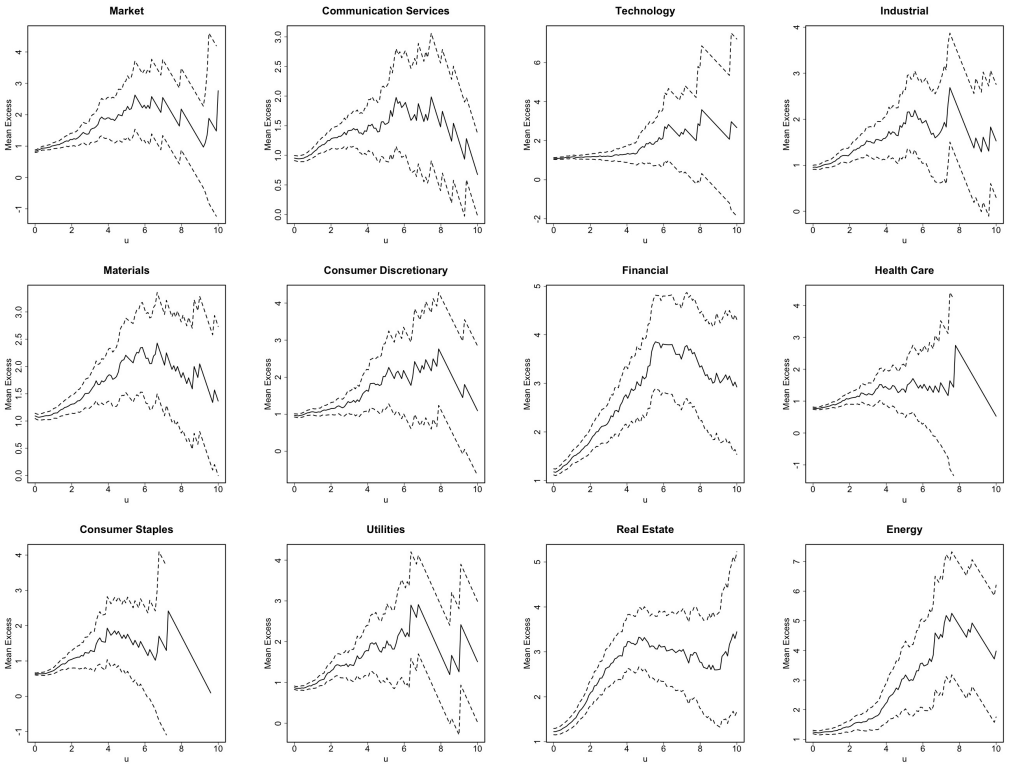

Figure 3.1: Mean residual life plots of each sector. The x-axis has the same scale as actual losses.

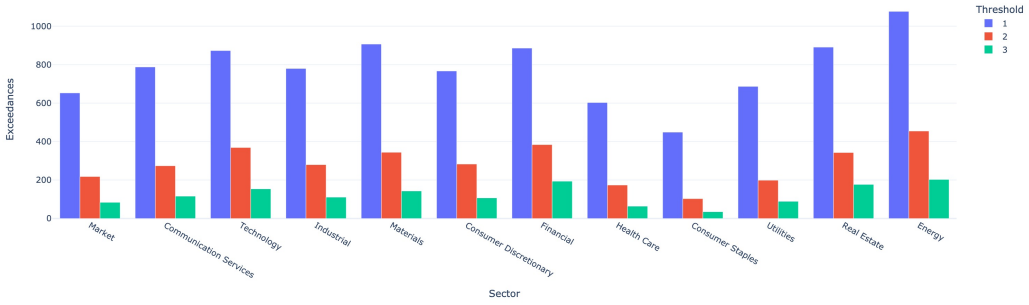

Figure 3.2: Number of threshold exceedances by sector.

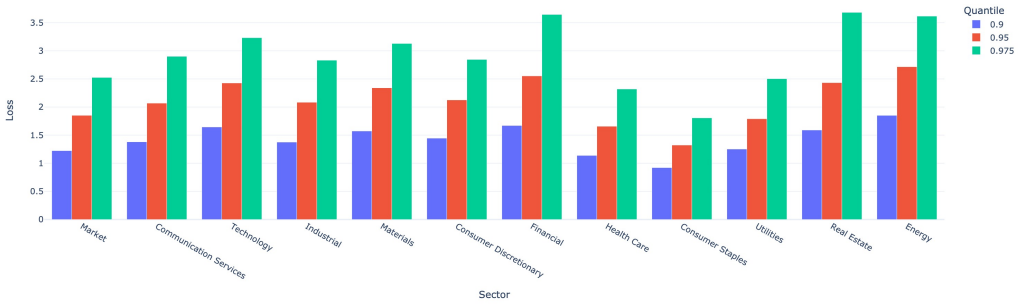

Figure 3.3: Sector loss quantiles estimates at levels $0.9, 0.95$ and 0.975.

The second method also involves estimating a GPD at different thresholds and plotting the parameter estimates against the threshold. The shape parameter ξ should be approximately constant and the scale parameter σ should be linear in u based on the definition of the GPD. To simplify the visualization, we look at $\sigma_u^* = \sigma_u - \xi_u u$ instead of σ_u which should be constant with respect to u. Consequently, an appropriate threshold is the lowest u above which the estimated ξ and σ_u^* remain constant (Coles 2013).

The two methods suggest using the threshold $u = 2$. However, if we look at the quantiles by sector, we may have too few observations for Consumer Staples and Health Care, as shown in Figure 3.3. Coles (2013) presents an example using the Dow Jones index where the selected threshold is also $u = 2$, but since we have multiple sectors behaving differently, we may have to do some scaling or use different thresholds. This is an important issue to solve for interpretation in terms of actual losses and not quantiles, since losing say 5% of value in any sector should be considered the same to the investor.

3.3 Bivariate Analysis

3.3.1 Definitions

Two approaches in extreme value analysis are used in the next part of the research. The first is the **Component-Wise Block Maxima** approach. We compute the maximum loss over a period of one month, which corresponds to approximately $m = 25$ observations or working days. This gives $n = 202$

total observations. For a chosen pair of variables X and Y, we define

$$M_j = (\max_{i \in B_j} X_i \ , \ \max_{i \in B_j} Y_i \ ), \quad \text{with } B_j = \{(j-1)m + 1, \ldots, jm\}.$$

Despite being a wasteful approach since most of the data information is lost, this can be used in the non-parametric estimation of dependence functions, for example. The second approach is to work with **Threshold Exceedances**. For a chosen pair of variables X and Y, we define $Z = (X, Y)$. The difficulty here is to define what we consider as extreme observations for Z. For a threshold u, an observation can be labeled as extreme if $\min(X, Y) > u$, $\max(X, Y) > u$ or $X + Y > u$ for example. This approach will be used in most cases.

3.3.2 Fréchet Scale Transformation

We use two approaches to transform a variable X to the Fréchet scale, where the transformation is

$$\tilde{X} = \frac{-1}{\log \hat{F}_X(X)}.$$

The first approach is to estimate the distribution empirically, i.e. using a rank-based empirical transformation. The second is to fit a GPD on the exceedances of a specified threshold u. We denote its parameters ξ, the shape, σ, the scale and ζ, the rate of threshold exceedance, i.e. the number of observations above u divided by the total number of observations. The transformation

$$\tilde{X} = \frac{-1}{\log\left(1 - \zeta\left[1 + \frac{\xi(X-u)}{\sigma}\right]^{-1/\xi}\right)},$$

follows approximately a standard Fréchet distribution for $X > u$ (Coles 2013). For $X < u$, we use a rank-based empirical transformation.

We can apply these transformations to the marginals of each pair of sectors and get Figure 3.4 and Figure 3.5. We use the function **gpd.fit** (Heffernan and Alec G. Stephenson 2018) for the GPD fitting. We use a color mapping to describe the strength of asymptotic dependence. We estimate the dependence measure χ for each pair of sectors and map it to a color, which goes from red, the weakest dependence, to green, the strongest. χ is called the extremal correlation, and is defined in the next section.

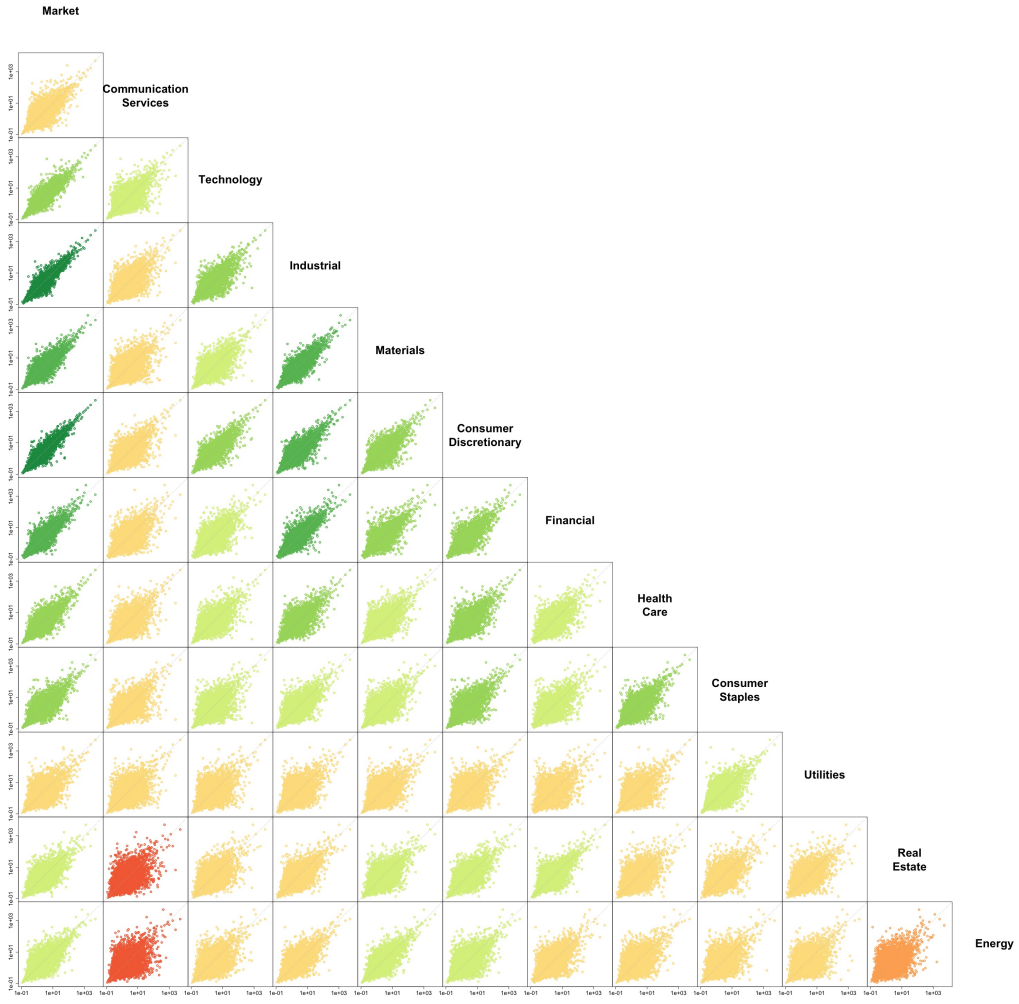

Figure 3.4: Sector pairs after empirical (rank based) transformation to the Fréchet scale. The color scale depends on the estimated extremal correlations and goes from red, the weakest dependence, to green, the strongest. Both axes are on a log scale.

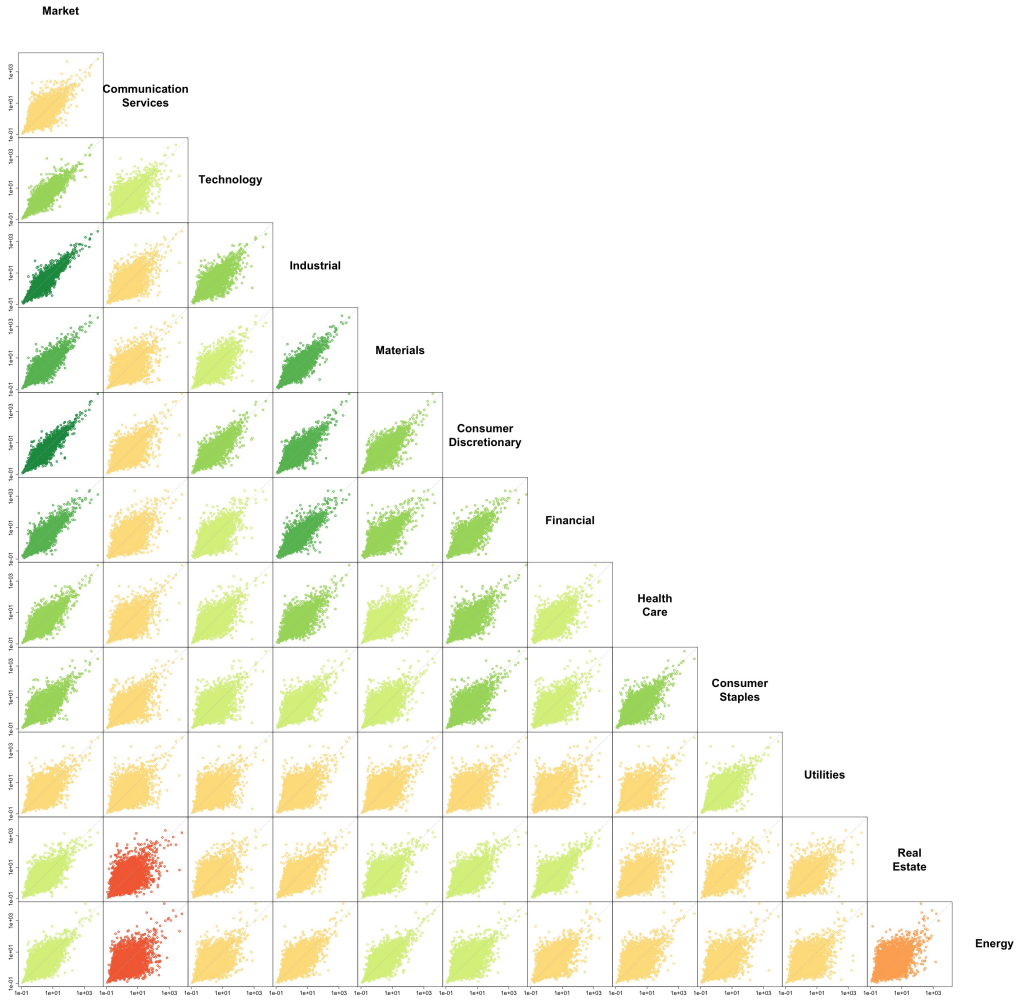

Figure 3.5: Sector pairs after GPD transformation to the Fréchet scale. The color scale depends on the estimated extremal correlations and goes from red, the weakest dependence, to green, the strongest. Both axes are on a log scale.

There is dependence between all sectors but its strength changes depending on the pairs. The weakest dependencies are between the Communication Services, Real Estate and Energy sectors. The strongest dependencies are between the Industrial, Materials, Consumer Discretionary and Financial sectors. The strength of dependence between the sectors and the whole market is very strong for all sectors except for Utilities and Communication Services. It is slightly weak for Real Estate and Energy sectors.

There is a symmetry in the plots, with some exceptions in the extremes for some pairs of sectors. The symmetry also varies depending on which transformation we apply, since the GPD modifies the extreme observations above the threshold. Significant losses in the market implies important losses in dark green sectors, though not for all sectors. The number of points above and below the diagonal (at extreme levels) depends on the transformation but it looks like there can be asymmetry. This would imply that few sectors out/under perform the market during huge losses, except Industrial and Consumer Discretionary which are strongly dependent on the market. However, these interpretations are in terms of standardized losses. In the next section, we will define an extremal asymmetry coefficient to showcase these results.

Finally, it appears that the Industrial sector is the building block of the market since the scatter plot is centered around the diagonal, the extremal dependence is quite high for all other sectors and represents the center of the most dependent cluster, as explained in the next section.

3.3.3 Logistic Model

We fit the bivariate extreme value logistic distribution logistic to threshold exceedances for pairs of sectors with threshold $u = 2$ for all sectors. In Figure 3.6 we plot the density provided by the fitted model for each pair of sectors and the estimates of quantile curves at the lower tail probabilities 0.8, 0.85, 0.9 and 0.95. The lines at extreme levels group the combinations of the two indicators having the same risk level. This shows that losses can be asymmetric at extreme levels, as we saw with the previous figures. In particular we can see the sectors that outperform (Consumer Staples) and underperform (Financial, Real Estate, Energy) the market during corrections, i.e., large losses.

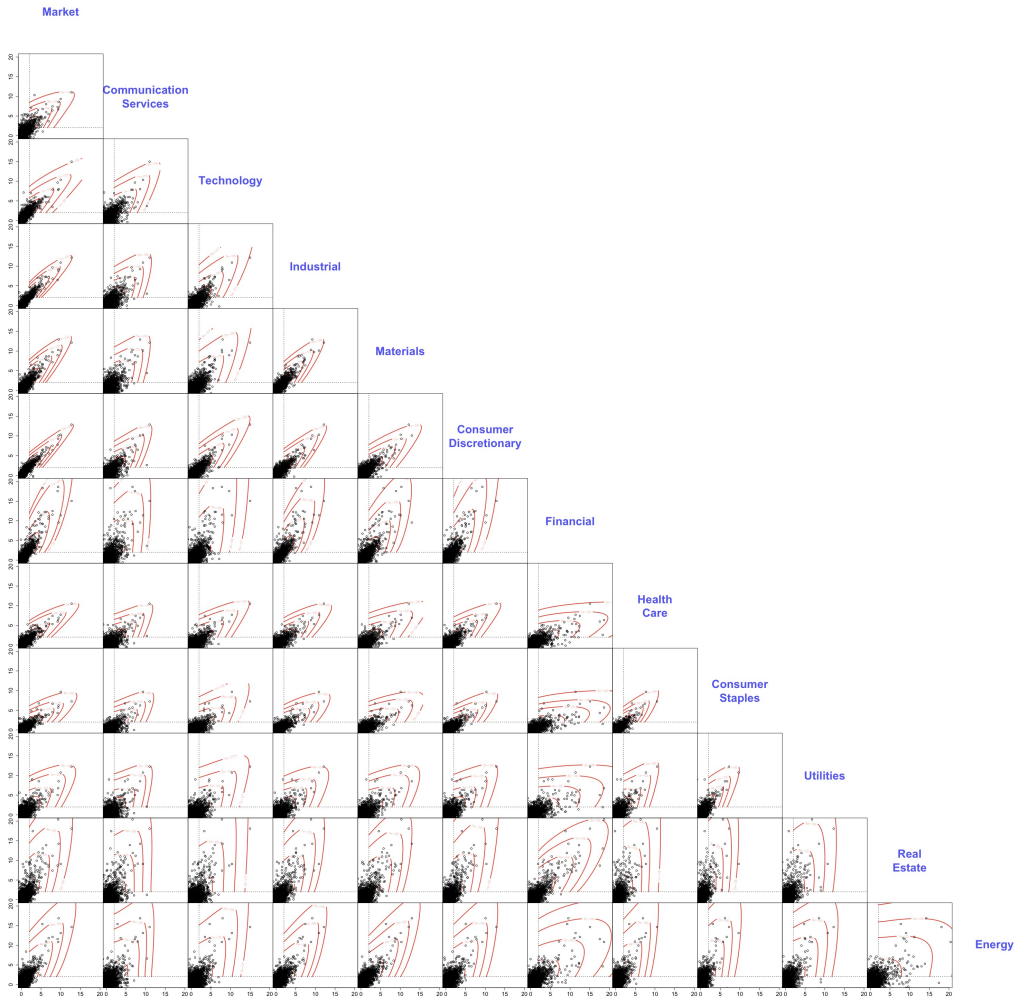

Figure 3.6: Estimates of the bivariate quantile curves (red curves) based on the bivariate extreme value logistic model. Dashed lines represent the thresholds $u = 2$. Only positive losses are shown.

3.4 Extremal Dependence

We present several extremal dependence measures applied on pairs of sectors.

3.4.1 Extremal Correlation Chi

Let X and Y two variables with marginal distribution functions F_X and F_Y. We define the extremal correlation as

$$\chi = \lim_{q \to 1} \mathbb{P}\{F_Y(Y) > q \mid F_X(X) > q\}.$$

This measure describes the tendency for one variable to be large conditional on a second being large. If $\chi = 0$ then X and Y are said to be asymptotically independent. If $\chi > 0$ then X and Y are said to be asymptotically dependent.

The χ plot is constructed by empirically estimating $\chi(q)$ and $\bar{\chi}(q)$ for different levels q, where

$$\chi(q) = 2 - \frac{\log \mathbb{P}\{F_X(X) < q, F_Y(Y) < q\}}{\log q},$$

$$\bar{\chi}(q) = \frac{2 \log(1-q)}{\log \mathbb{P}\{F_X(X) > q, F_Y(Y) > q\}} - 1.$$

We are interested in the limits of $\chi(q)$ and $\bar{\chi}(q)$ when q goes to 1, which summarize the strength of dependence between the variables (Coles 2013). We can show that $\chi(q)$ converges to χ, and we denote $\bar{\chi}$ the limit of $\bar{\chi}(q)$. If $\bar{\chi} = 1$, we have asymptotic dependence and we use χ as a measure of dependence within the class of asymptotic dependence. Otherwise, we use $\bar{\chi}$ as a measure of dependence within the class of asymptotic independence.

Figure 3.7 shows the plots of χ and $\bar{\chi}$ for all pairs of sectors, made using the R function **chiplot** (A. G. Stephenson 2002). Both axes of the plots have the range $[0, 1]$, and the confidence intervals are computed based on normality assumptions and are at level 95%. We are interested in the asymptotic behavior of the two lines for $u \to 1$. Most of the plots have the same shape, where $\bar{\chi}$ converges to 1 and χ decreases slightly and converges to around 0.5. The interpretation is not straightforward because of the large variance of estimators as seen in the confidence intervals but it supports asymptotic dependence. There is a difference in the dependence structures depending on the sectors.

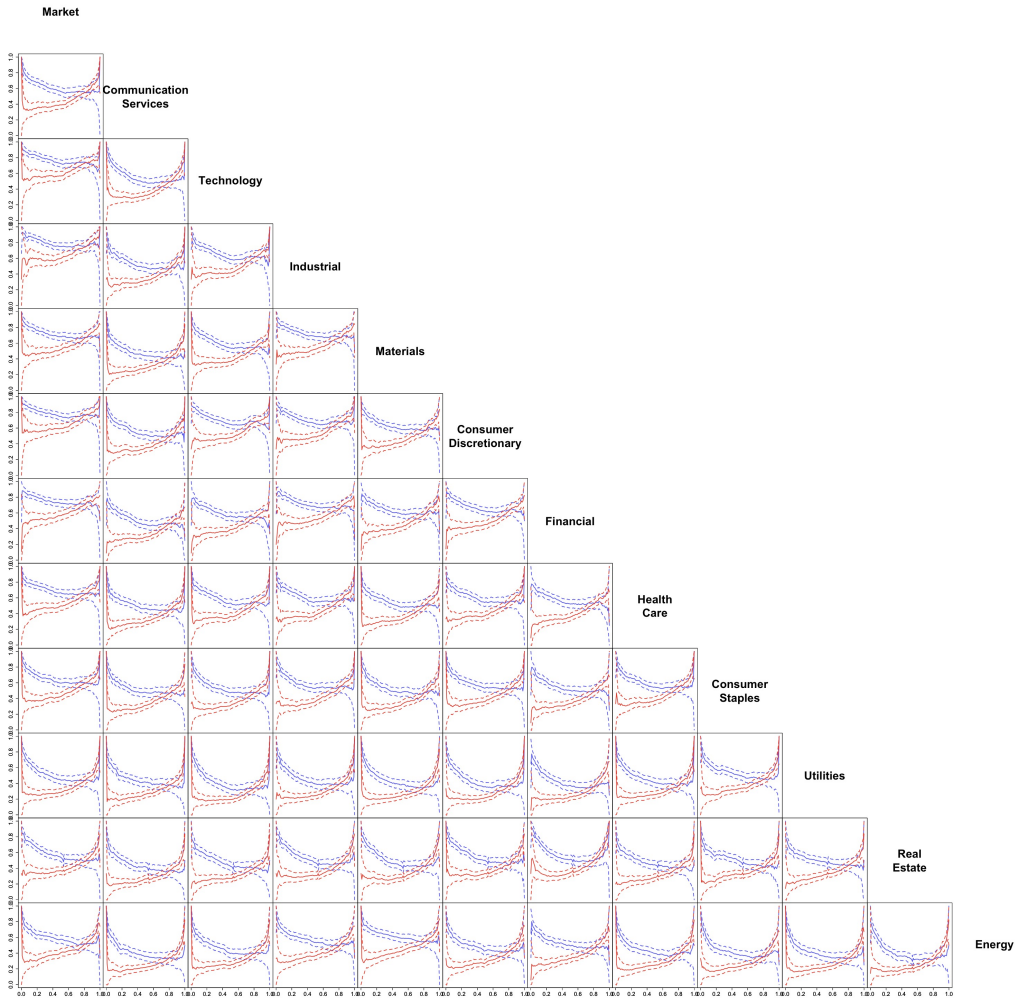

Figure 3.7: Empirical estimates of extremal correlations $\chi(q)$ (blue) and $\bar{\chi}(q)$ (red) and their point-wise 95% confidence intervals (dashed lines). Both x-axis and y-axis range is [0,1].

To confirm these results, we use a different approach by computing another empirical estimate of χ. Our motivation here is to better understand relationships between the biggest losses by month, which leads us to investigate monthly extremes. We previously defined the variable as M_j for each month $j = 1, \ldots, m$ such that $M_j = (M_j^1, M_j^2)$. We denote the estimated marginal empirical distribution functions by $\hat{G}_1$ and $\hat{G}_2$. The max-stability assumption implies a deterministic relationship between the F-Madogram of the marginals and the extremal correlation χ (Huang et al. 2019). In this case, the empirical estimator of χ is

$$\hat{\chi} = 2 - \frac{1 + 2\hat{\nu}}{1 - 2\hat{\nu}},$$

where

$$\hat{\nu} = \frac{1}{2m} \sum_{j=1}^{m} |\hat{G}_1(M_j^1) - \hat{G}_2(M_j^2)|.$$

We summarize the estimates of the dependence of each pair in a matrix, as seen in Figure 3.8. Using this matrix, we can build an undirected and weighted graph, where the nodes represent the eleven sectors. Two sectors are connected by an edge if the estimated χ isn't zero, which implies dependence between the sectors. The weights of each edge is equal to the corresponding χ estimate, which describes the strength of dependence. We draw the network in Figure 3.9. The chosen colors correspond to possible clusters of sectors based on the strength of asymptotic dependence of the components. First, the conclusions from the bivariate analysis are coherent with what is observed in the matrix, in particular the dependence between the market and its components and the asymptotic dependence between the sectors. Looking at the network, the four sectors in red are the most dependent with $\chi > 0.7$. The two sectors in green are also quite dependent and share a connection with the first cluster. Technology is also dependent on the Industrial and Consumer Discretionary sectors of the red cluster, with $\chi > 0.6$. The remaining sectors share few significant connections with the others and represent clusters by themselves. We end up with seven clusters or subgroups of sectors that are defined by the colors in Figure 3.9. These results can be useful in reducing the number of faces to consider when working with exponent measures with the objective of finding concomitant extremes in the multivariate case (Engelke and Ivanovs 2021). It decreases the maximum number of faces from $2^{11} - 1 = 2047$ to $2^7 - 1 = 127$, which is a significant drop.

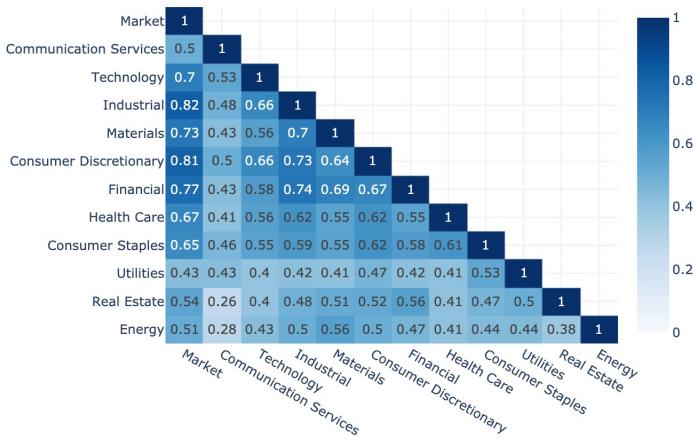

Figure 3.8: Empirical estimate of the extremal correlation χ matrix using monthly maximum losses computed pair by pair.

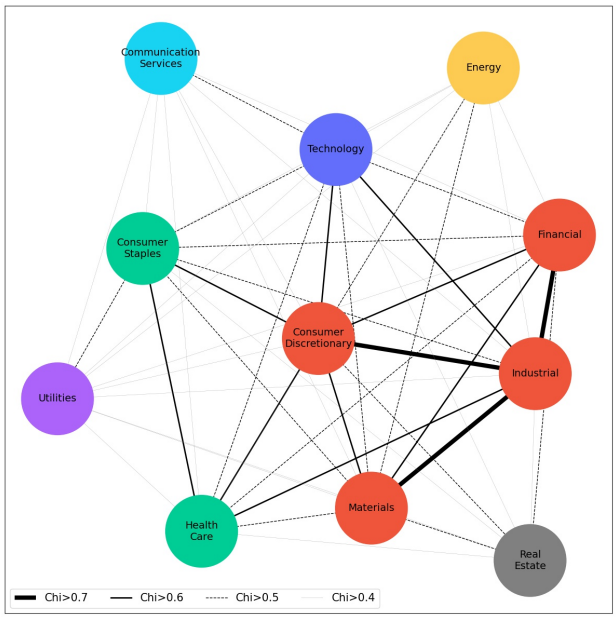

Figure 3.9: Plot of χ network based on the χ matrix. Nodes represent sectors. Edges correspond to existing dependencies. Weights are equal to the corresponding χ. Colors define clusters.

The colors in Figure 3.9 are selected as follows. For the threshold $\chi > 0.7$, the four sectors in red are connected, where the Industrial sector plays the center. When decreasing the threshold to $0.6 < \chi < 0.7$, new connections, i.e. edges, are drawn and we select new colors. Technology is only connected to sectors in the previous cluster, so we choose a new color for it. However, Consumer Staples and Health Care share a new connection and are not in the red cluster, so we choose a new color for both. Now we have three clusters. The remaining sectors share connections with the previous clusters but not with each other, for $0.5 < \chi < 0.6$, so each sector takes a new color.

3.4.2 Coefficient of Tail Dependence Eta

A second tail dependence coefficient called η is constructed as follows. If Z_1 and Z_2 have unit Fréchet marginal distributions then under broad conditions (Davison 2021) we have

$$\mathbb{P}(Z_1 > t, Z_2 > t) \sim L(t)\mathbb{P}(Z_1 > t)^{\frac{1}{\eta}}, \quad t \to \infty,$$

where the coefficient of tail dependence $\eta \in (0, 1]$ and L is slowly varying at ∞, i.e. $L(ct)/L(c) \to 1$, as $c \to \infty$, and $t > 0$. One can also show that

$$\bar{\chi} = 2\eta - 1,$$

so if the variables are asymptotically dependent, i.e. $\bar{\chi} \sim 1$, then $\eta = 1$, and if they are independent, i.e. $\bar{\chi} \sim 0$, then $\eta = 1/2$, and inversely. There are different approaches to estimating η. We present a simple approximation that uses the GPD. Let $T = \min(\tilde{X}, \tilde{Y})$, where $\tilde{X}$ and $\tilde{Y}$ are the transformations of the two original variables X and Y to the Fréchet scale. For a selected threshold u, if we fit a GPD to the exceedances of T then the shape parameter provides an estimate of η. We present a sketch of the proof (Davison 2021). As

$$\mathbb{P}(T > t) = \mathbb{P}(\tilde{X} > t, \tilde{Y} > t),$$

we have for a large enough t and u that

$$\mathbb{P}(T > t + u \mid T > u) \sim \frac{L(z + u)(t + u)^{-1/\eta}}{L(u)u^{-1/\eta}} \sim (1 + t/u)^{-1/\eta},$$

and that can be written as

$$\mathbb{P}(T - u > t \mid T > u) \sim (1 + \eta\frac{t}{\tilde{\sigma}})^{-1/\eta}, \quad u, t > 0, \quad \tilde{\sigma} = \eta u.$$

The survivor function of $T - u$ conditional on $T > u$ has the form of the GPD, so if we fit a GPD to the threshold exceedances then the shape parameter ξ provides an approximation of the tail dependence coefficient η. However, η depends on the selected threshold so we perform the estimation at increasing values of u and look at the limit.

We use the function **tcplot** (A. G. Stephenson 2002) to plot the estimated shape parameter against the probability of exceeding the corresponding threshold and summarize the results for the different pairs of sectors in Figure 3.10. The x-axis range is $[0, 1]$ and the y-axis is $[0.5, 1.2]$. The red dotted line is $y = 1$ corresponds to asymptotic dependence. The η plots confirm the results from the χ plots. Clearly there is a dependence between all pairs of sectors but its strength varies. Most of the plots are linearly increasing to values around 1. However, there are some pairs where we observe a significant drop at the limit, such as Utilities/Real Estate. The confidence intervals are quite large, making the interpretation harder, as seen with χ plots.

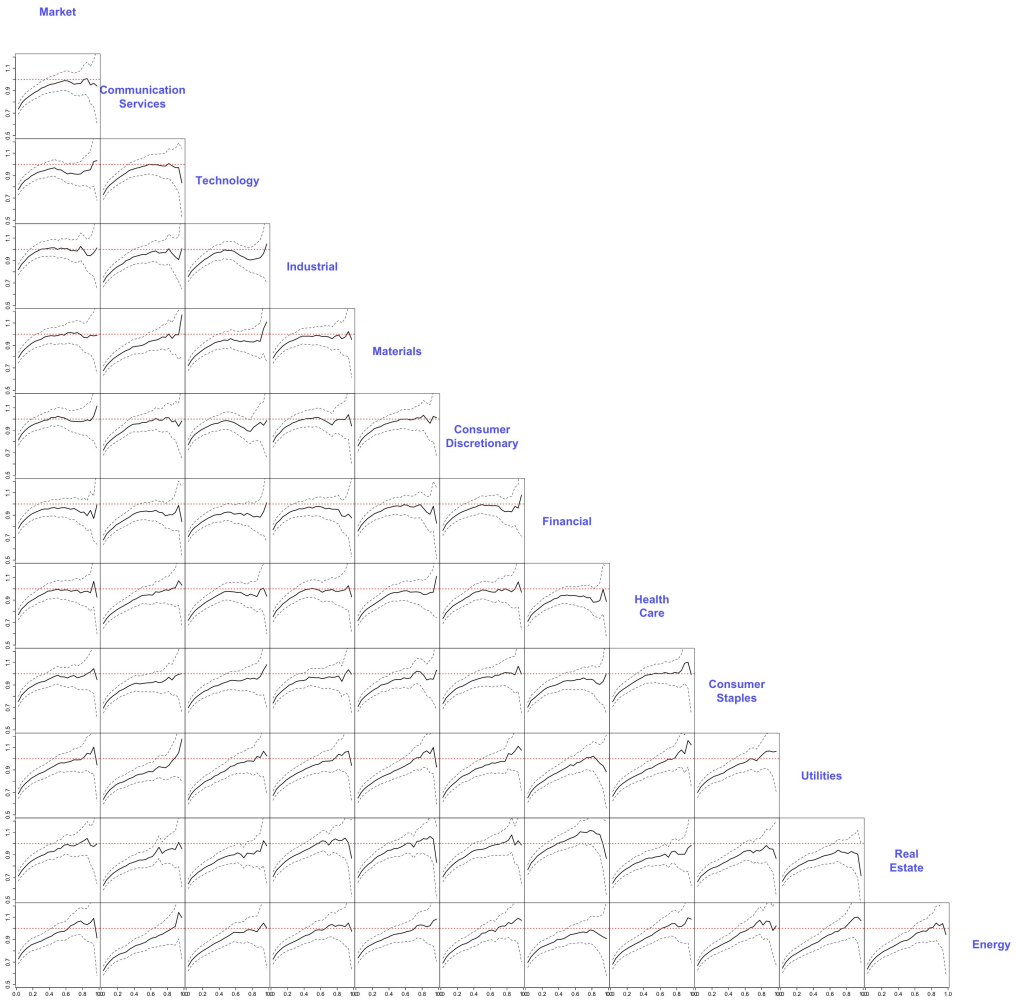

Figure 3.10: GPD estimate of the coefficient of tail dependence $\eta(u)$. The dashed red line at $y = 1$ represents asymptotic dependence. The x-axis range is [0,1] and the y-axis is [0.5,1.2].

3.4.3 Pickands Dependence Function

A richer summary of the dependence structure between two variables is provided by Pickands' dependence function A, which satisfies

- $\max(t, t-1) \leq A(t) \leq 1, \quad t \in [0,1]$,
- $A(t) = 1$ for independent data,
- $A(t) = \max(t, t-1)$ for perfectly dependent data,
- A is convex.

The main difficulties with this measure are the estimation process, the transformation of the variables and the selection of the observations used for the estimation. The estimation can be empirical, or based on a parametric model such as the logistic or Hüsler–Reiss. The transformation can involve the maximum approach used in GEV models, or the re-scaling of the variables. Also, in most cases we should have two variables on the Fréchet scale and we need to select which sets of points are used to estimate A. Since we are interested in the extremes, we need to define threshold exceedances for the bivariate case. Three main sets commonly used are

- $\mathcal{A} = \{(X,Y) \ : \ \min(X,Y) > q\}$,
- $\mathcal{A}' = \{(X,Y) \ : \ \max(X,Y) > q\}$,
- $\mathcal{A}'' = \{(X,Y) \ : \ X + Y > q\}$,

where X, Y are the two variables on the Fréchet scale and $q \in [0,1]$ is the level or threshold. In this section, we will look at the different approaches proposed and compare the results. The function **abvnonpar** (A. G. Stephenson 2002) provides a non-parametric estimate for the dependence function $A(t)$ of the bivariate extreme value distribution. An empirical transformation of the two marginals is performed and the new variables are denoted y_1 and y_2. The Pickands dependence function estimate is

$$A_q(w) = \frac{n}{\sum_{i=1}^{n} \min(y_{i1}/w, y_{i2}/(1-w))},$$

but a modification of A is implemented to satisfy the constraints above, viz

$$\tilde{A}_q(w) = \min\{1, \max(A_q(w), w, 1-w)\}.$$

First, we use the component-wise block maxima approach. In Figure 3.11 we also look at other estimates of the dependence function such as the Deheuvels, Hall–Tajvidi and Capéraà–Fougères–Genest (CFG) estimations, which are all based on marginal adjustments (Capéraà, Fougères, and Genest 1997). The only dependence function that is always convex is the latter, drawn in red, which estimate is

$$\log(A_q(w)) = \frac{1}{n}\left\{\sum_{i=1}^{n}\log(\max[(1-w)y_{i1}, wy_{i1}]) - (1-w)\sum_{i=1}^{n}y_{i1} - w\sum_{i=1}^{n}y_{i2}\right\}.$$

One can use the convex minorant to make the estimators convex but this hasn't been done in this example. The strength of dependence is close to what the χ plots and χ network suggested. The weakest dependence is between Communication Services and Energy/Real Estate. The strongest is between the sectors Industrial, Materials, Consumer Discretionary and Financial. Now we estimate the dependence function based on the threshold exceedances to see the effect of threshold selection. The results should stay coherent when increasing the threshold at appropriate levels. For the set of points $\mathcal{A}$, we use the same estimation procedure as before, but using the observations exceeding a threshold, then plotting the corresponding CFG dependence functions. The results are coherent with the previous conclusions and do not depend on the threshold, as seen in Figure 3.12. As a remark, we select the exceedances using actual losses but the estimation involves an empirical transformation. However, for the set $\mathcal{A}'$, the dependence structure is significantly different. Most of the pairs are less dependent and the functions are no longer convex. The estimation procedure is not appropriate for this set of observations because the convergence to the extreme value limit might be doubted. For the last set $\mathcal{A}''$, we use a different estimation approach based on a weighted empirical likelihood of the angular distribution function. The two variables are transformed to have unit Fréchet margins, then we use the angular measure to estimate the dependence function. The code is not released yet and was kindly shared by a PhD student and colleague Alouini (2022). The results seem coherent with the previous analysis too and the threshold has no effect on the estimation.

In conclusion, all the dependence measures used in this section induce the same conclusions, so Figure 3.9 is a good summary of the structure, that we will model in the following parts of the report.

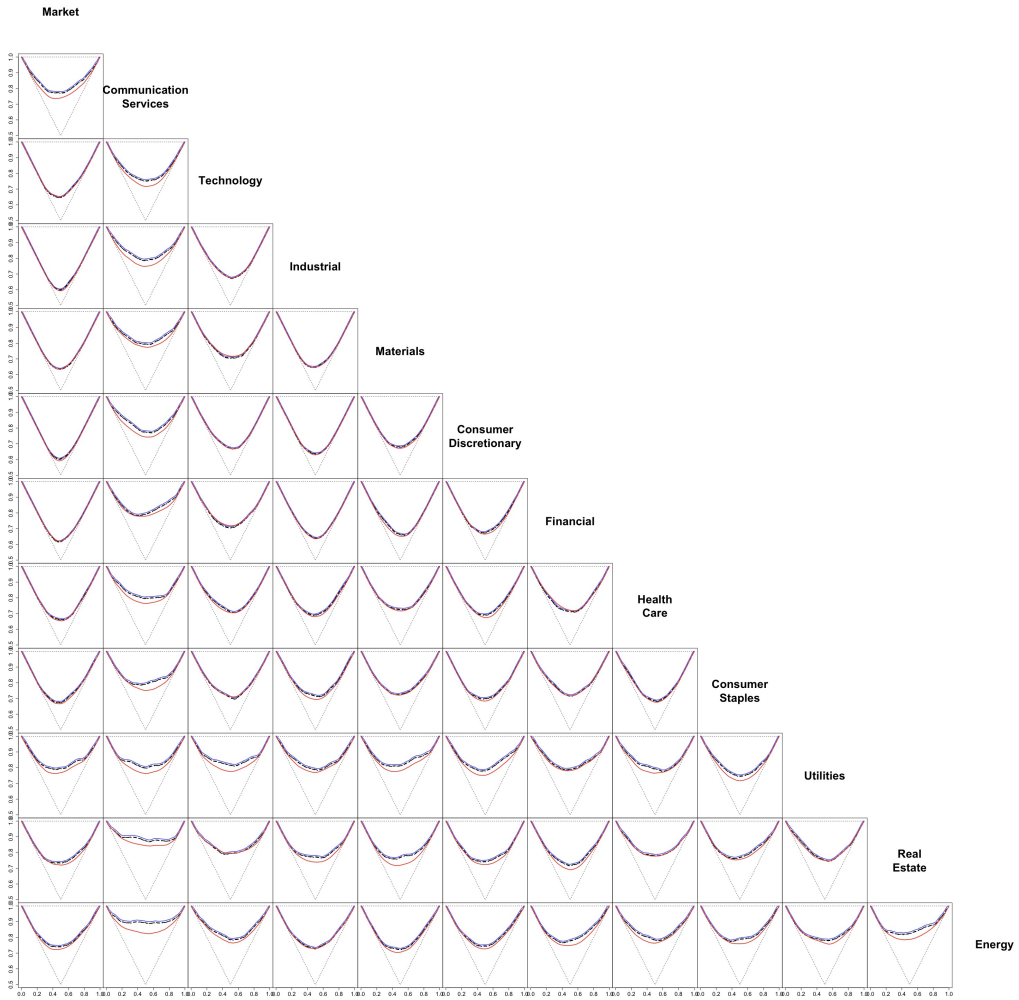

Figure 3.11: Non-parametric estimates of dependence functions. Pickands (blue), Deheuvels (dashed), Hall–Tajvidi (dashed) and Capéraà–Fougères–Genest (red). Component-wise block maxima approach.

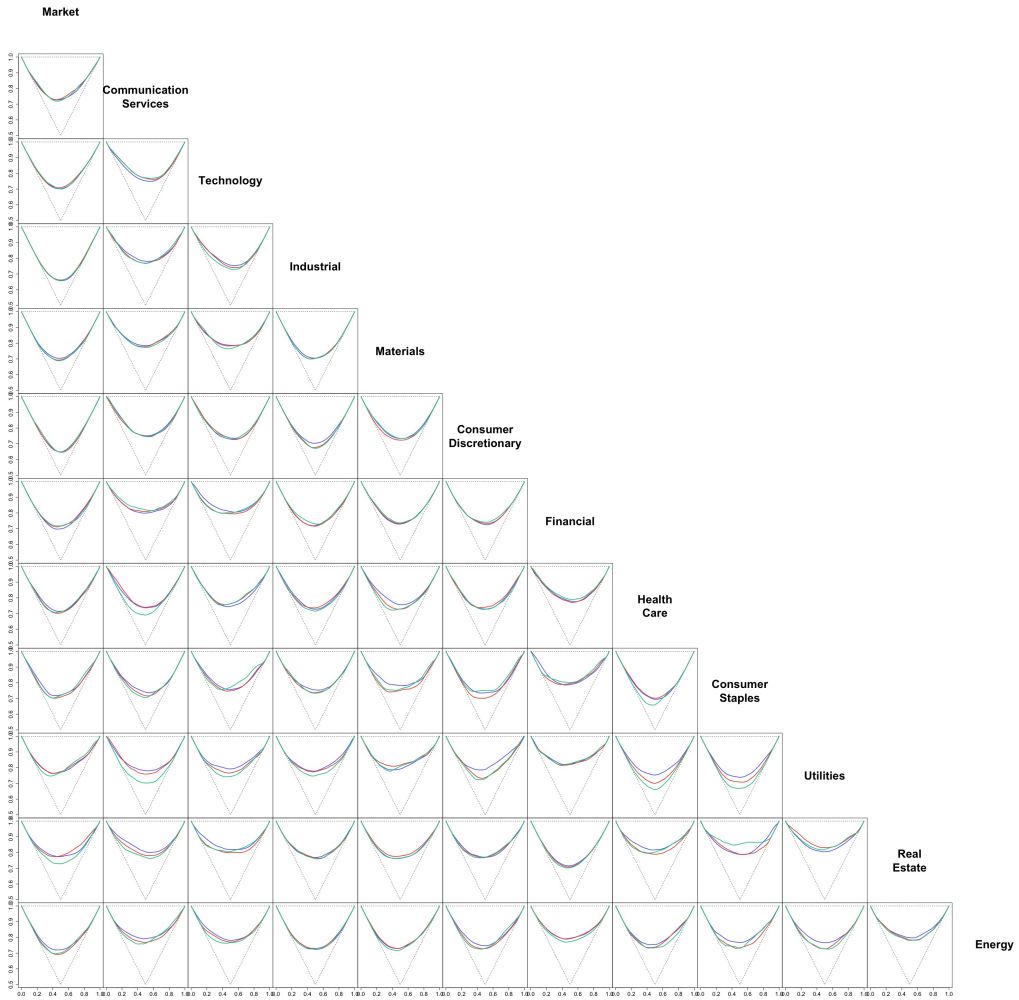

Figure 3.12: Non-parametric estimates of dependence functions at different thresholds: $u = 1.5$ (blue), $u = 2$ (red), $u = 3$ (green). The Capéraà–Fougères–Genes variant is used.

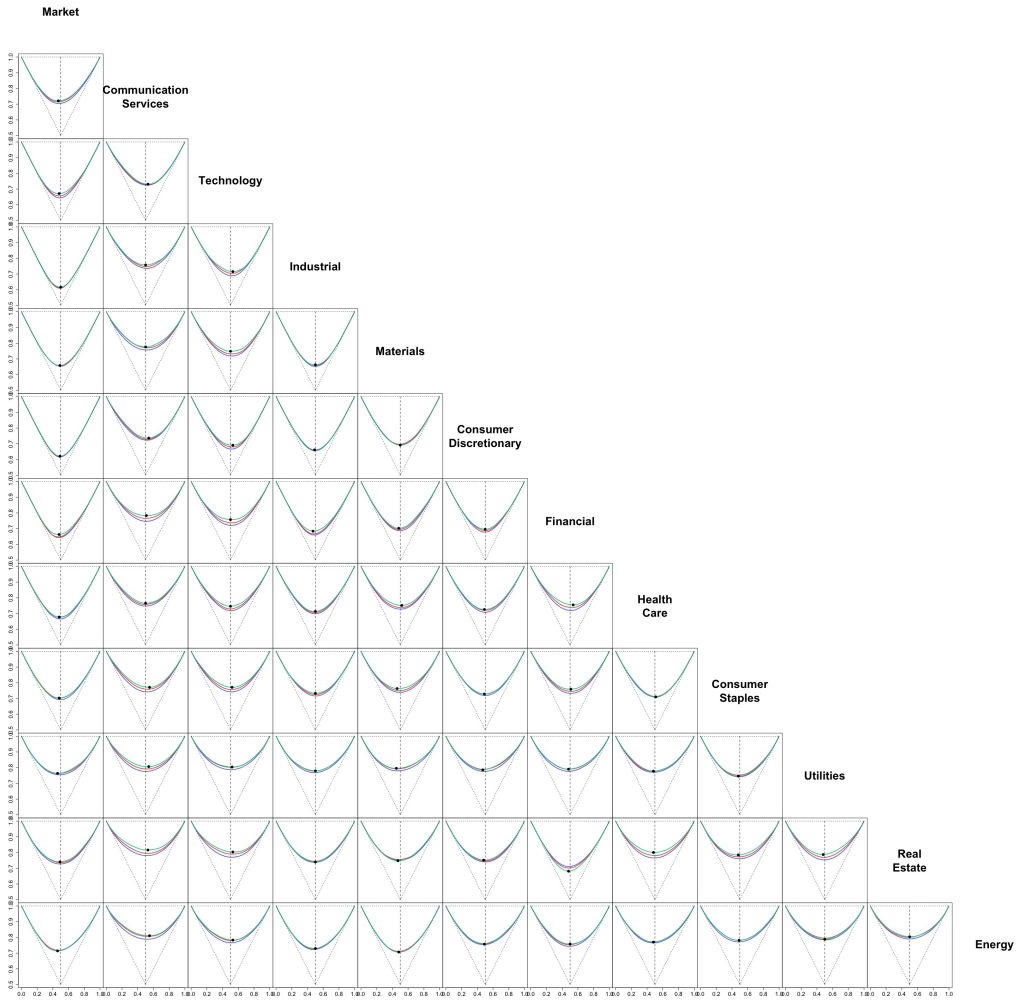

Figure 3.13: Weighted empirical likelihood estimation of Pickands' dependence function at different levels: $q = 0.8$ (blue), $q = 0.9$ (red), $q = 0.95$ (green). The point is the minimum at level 0.95 (green curve).

3.5 Extremal Asymmetry

We previously made remarks related to the asymmetry in the extreme observations when we plotted all the sector pairs after transformation to the Fréchet scale. In this section, we quantify the level of asymmetry in the dependence structure. We introduce the coefficient of extremal asymmetry that describes the relative tendency of one variable to be larger than the other, given that both are extreme, as explained by Semadeni (2020). It is constructed as follows.

Given a level $q \in [0, 1]$ and two variables X and Y, we empirically transform the two variables and write $U = \hat{F}_X(X)$ and $V = \hat{F}_Y(Y)$. Then we define

$$\rho(q) = \frac{\mathbb{P}(V > U | U > q, V > q) - \mathbb{P}(V < U | U > q, V > q)}{\mathbb{P}(V > U | U > q, V > q) + \mathbb{P}(V < U | U > q, V > q)},$$

and define the coefficient of extremal asymmetry as

$$\rho = \lim_{q \to 1} \rho(q).$$

We have that $\rho \in [-1, 1]$. If $\rho > 0$ then V tends to be asymptotically larger than U and inversely for $\rho < 0$. If $\rho = 0$ then the variables are asymptotically symmetric. The confidence intervals can be computed using bootstrapping, where the intervals are supposed to be normal.

Figure 3.14 summarizes the results for the different pairs of sectors. The plots are mostly flat for $q < 0.8$, then, depending on the pair, there is a peak or drop at the limit. For example, the pair Market/Communication Services has a limit, $\rho \sim 0.5$ which means the Communication Services sector tends to have smaller losses than the whole market during corrections. The only sector that underperforms the market during large losses is the Consumer Discretionary sector. We also see that the Industrial sector is quite symmetric with the market. These results are in terms of quantiles or standardized losses. Since they are quite different depending on the sector, we are trying to find another approach to simplify the interpretation in terms of actual losses.

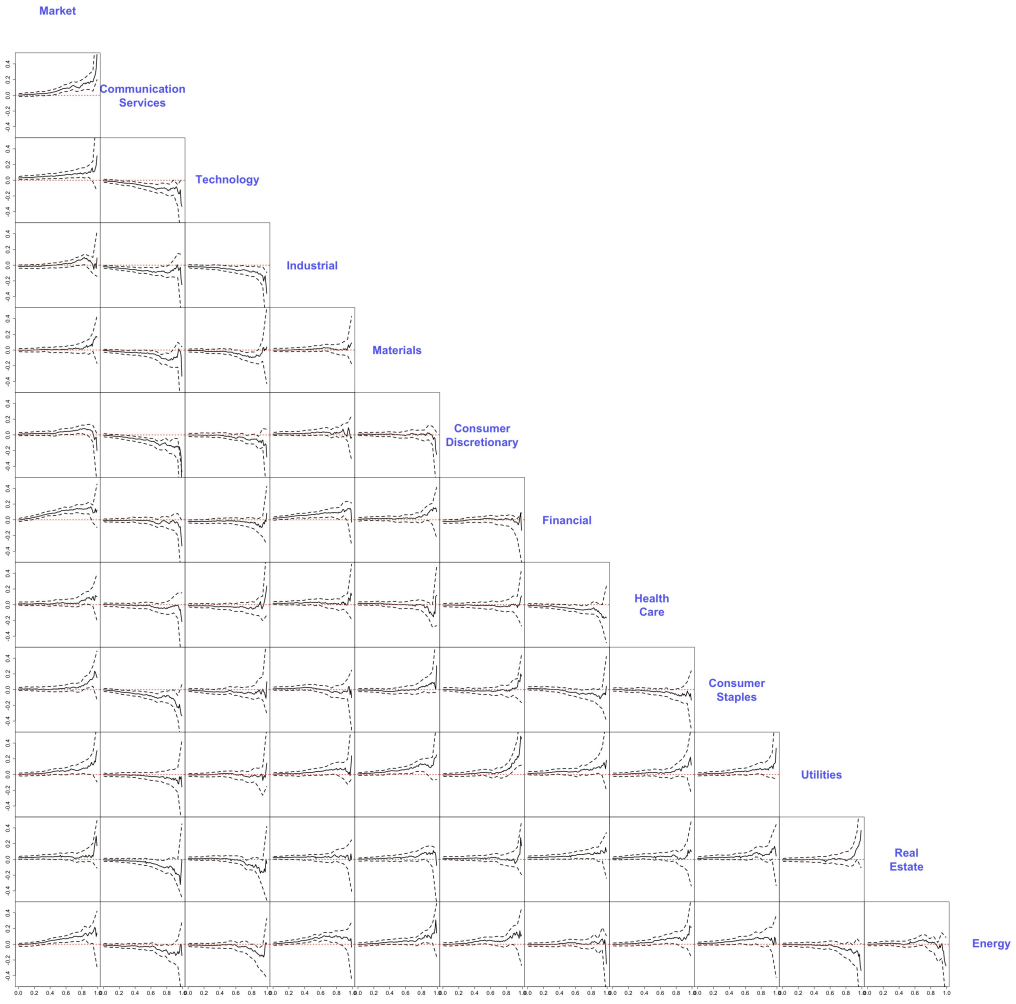

Figure 3.14: Empirical estimates of extremal asymmetry $\rho(q)$. The dashed red line at $y = 0$ represents asymptotic symmetry. The x-axis range is [0,1] and the y-axis is [-0.5,0.5]. If $\rho > 0$ then the sector above the plot tends to asymptotically have larger losses than the sector on the right.

Another way to characterize extremal asymmetry is to use the previous Pickands dependence function in Figure 3.13 (Semadeni 2020). Suppose (X, Y) is a bivariate random vector following an extreme value distribution G with Pickands dependence function A and unit Fréchet margins. If the exponent distribution associated with G is differentiable then we have

$$\rho = \frac{A'(1/2)}{2 - 2A(1/2)},$$

where A' is the derivative of A. In this case, the null hypothesis

$$H_0 : X \text{ and } Y \text{ are asymptotically symmetric}$$

is equivalent to

$$\rho = 0 \iff A'(1/2) = 0.$$

We can look at the minimum achieved by the estimated Pickands dependence function, which is plotted as a dot in Figure 3.13. Three cases are possible: If the minimum is achieved at $1/2$ then we have symmetry. Otherwise, we have $\rho > 0$ if it is attained before $1/2$ and $\rho < 0$ after. All sectors, except Industrial with $\rho = 0$, tend to have asymptotically larger losses than the Market, i.e. $\rho > 0$. However the values are close to zero, and in the case of Consumer Discretionary, Figure 3.14 shows that ρ should be negative. The pair Financial/Health Care shows a negative ρ, so the Health Care sector tends to have asymptotically larger losses than the Financial sector. Figure 3.14 shows it is indeed negative, but the confidence interval contains zero. The dependence function or the extremal asymmetry plots should be used carefully. Both help detect asymmetry in standardized but not in the actual losses. For example, if one sector appears to be out-performing another one based on ρ and the two sectors have very different quantiles for their losses then it is possible that they are performing the same, or even the first sector is actually under-performing relative to the second, in terms of actual losses.

Next chapter provides a multivariate analysis of extreme losses in the S&P 500 sectors. We present a tail dependence matrix decomposition, that allows exploring extremal dependence in high dimensions. We use the tail dependence matrix to compute partial correlations, then describe the dependence between sectors as a Gaussian graphical model. We present an adaptation of k-means clustering to extremes. We fit a conditional multivariate extreme value model to the S&P 500 sectors, where the conditioning variable is the S&P 500 index.

4. Multivariate Analysis

4.1 Tail Dependence Decomposition

In the previous chapter, we used different approaches to quantify the dependence between two sectors then summarized the results as a matrix containing the information for all pairs. We are now interested in a different way to summarize and describe high-dimensional tail dependence, so we consider a decomposition that provides tools for exploring extremal dependence in high dimensions. This can be seen as an adaptation of principal component analysis (PCA) to extremes. We will employ the framework of regular variation and angular measures to create a positive semi-definite tail dependence matrix, then perform an eigen-decomposition as described by Cooley and Thibaud (2019). We first transform the data to meet the framework's assumptions.

Let $x_1, \ldots, x_n$ be the original observations for one particular index, i.e. sector or market. We apply the transformation $x_i \to y_i = \log(1 + \exp x_i)$ to bound the values away from 0, thereby meeting the lower-tail requirements. This transformation leaves the magnitudes of the large losses essentially unchanged. As mentioned above, the scales and shapes of the variables are quite different, so we apply a second transformation in order to have approximately the same scale and a tail index $\alpha = 2$, as required by the framework. It is suggested to use the transformation $y_i \to c_i^{-1/2} y_i^{\alpha_i/2}$, where c_i is the estimated scale and α_i the estimated tail index. However, we had trouble with this transformation, so we transformed the $y_1, \ldots, y_n$ to the Fréchet scale, using a rank-based empirical transformation, and then took the square root to meet the tail index assumption, i.e $y_i \to z_i = (-1/\log \hat{F}(y_i))^{1/2}$. By transforming the margins as above, we obtain a regularly varying random vector $Z \in RV_+^d(2)$, where $d = 12$ is the number of indices, that satisfies the

assumptions of the framework (Cooley and Thibaud 2019). We change the notations for simplicity, by denoting the observations of Z as $z_1, \ldots, z_n$, where $z_t \in \mathbb{R}^d,\ t = 1, \ldots, n$. Now we are interested in estimating the tail pairwise dependence matrix Σ_Z. We compute the polar representation, related to the angular measure, for each observation z_t and denote

$$r_t = ||z_t||_2\ , \quad w_t = z_t/r_t, \quad t = 1, \ldots, n.$$

We select a high threshold r_0 for the radial components, corresponding to the 95% quantile of r, then estimate the tail dependence matrix as

$$\hat{\Sigma}_Z = \frac{1}{n_{exc}} \hat{m} \hat{W}^T \hat{W},$$

where

$$n_{exc} = \sum_{t=1}^{n} \mathbb{I}(r_t > r_0)\ , \quad \hat{m} = \frac{r_0^2}{n} \sum_{t=1}^{n} \mathbb{I}(r_t > r_0),$$

and $\hat{W}$ is the matrix whose rows are the vectors w_t for which $r_t > r_0$. The estimated $\hat{\Sigma}_Z$ is positive semi-definite.

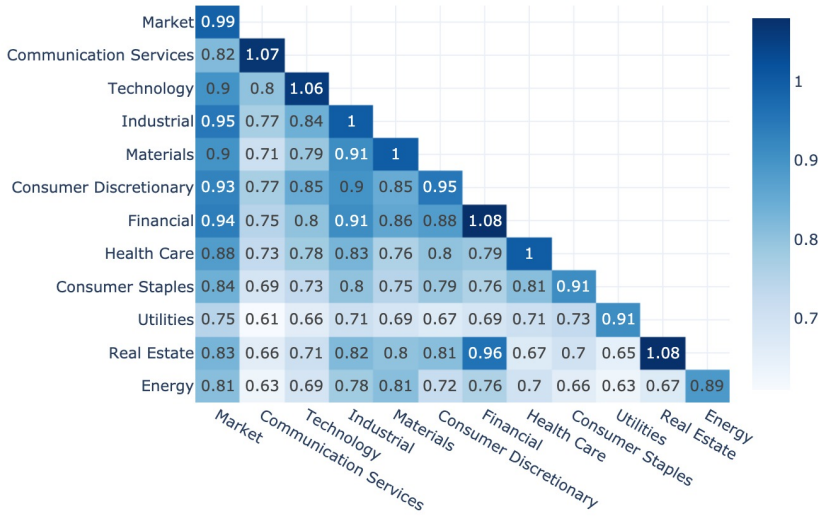

Figure 4.1: Empirical estimate of the tail pairwise dependence matrix based on the polar transformation of losses and using the observations with the largest 5% of Euclidean norms. Market is included.

Figure 4.1 shows the estimated matrix. The first column describes the dependence between the sectors and the market. The same four sectors as in Figure 3.9, plus Technology, again strongly depend on the market. Industrial has the strongest and Utilities the weakest dependence with the market. The same four sectors are very dependent with each other, as seen in the center block of the matrix. Now, we perform an eigen-decomposition and analyze the eigen-vectors, i.e. loadings, of the five largest eigen-values, as seen in Figure 4.2.

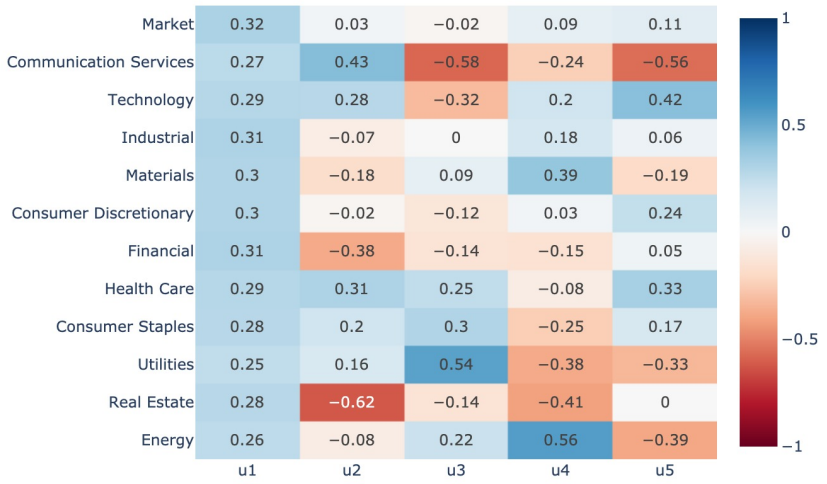

Figure 4.2: First five eigen-vectors, i.e. loadings, corresponding to the eigen-decomposition of the tail dependence matrix $\hat{\Sigma}_Z$. Market is included.

The eigen-vector u_1 gives an overall magnitude. u_2 contrasts the Financial and Real Estate sectors with the other sectors, in particular Communication Services, Healthcare and Technology. The third vector u_3 contrasts Utilities with Communication Services. We saw the same result in Figure 3.9, where these two sectors were the farthest from each other. We use the loadings to build the time series of principal components, i.e. scores, such that $v_i = \log(\exp z_i - 1) \cdot u_i^t$. We only keep positive losses for visualization purposes, as seen in Figure 4.3.

The three time series in Figure 4.3 clearly show financial crises, such as the 2008 global crisis and the 2020 COVID crisis. If we focus on the 2008 crisis, we can see that the large positive value of the first score arises from the large losses across the market, while the large negative value of the second score helps to mitigate the loss for Communication Services, Technology and Healthcare and accentuate losses for Real Estate and Financial. Also, the third score shows more volatility than the first two, especially during crises. Nearly the same thing happened during the COVID crisis, but with a much quicker rebound, and a peak in the second score. The first two scores achieved their highest positive value since 2002 during COVID. Figure 4.4 shows that the three principal component scores have very different scales, which means they don't imply the same effect on losses.

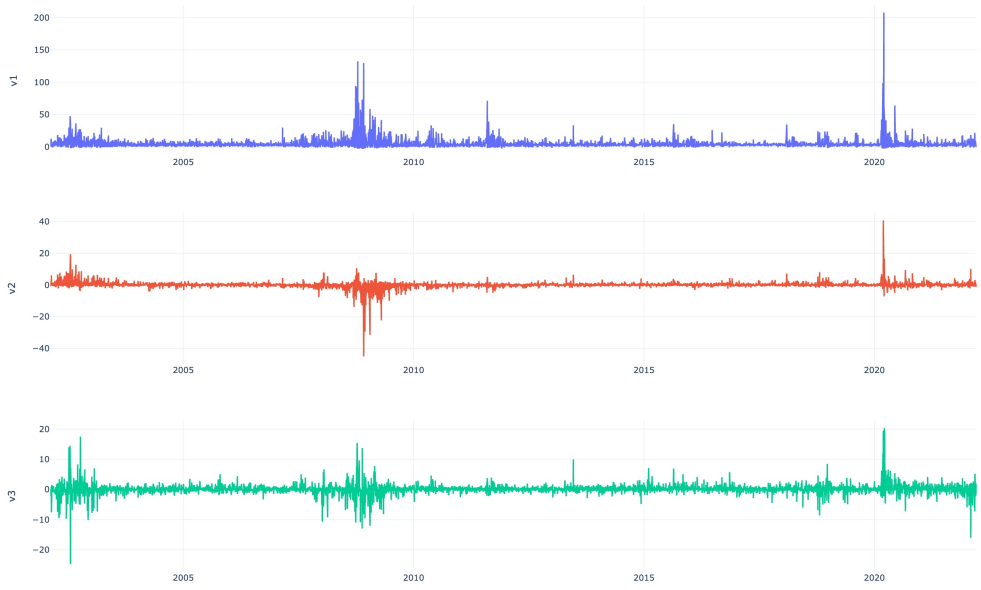

Figure 4.3: Time series of the first three principal components, i.e. scores. Blue is the first, red the second and green the third. The first score describes losses across all sectors. The second score contrasts losses between certain sectors. Market is included.

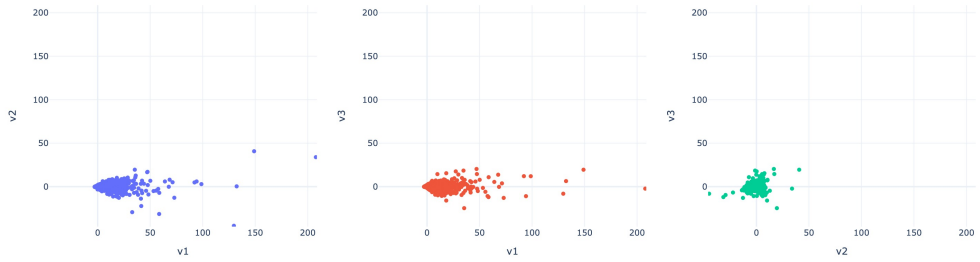

Figure 4.4: Scatterplots of pairs of principal components, i.e. scores. The scale of the second and third is much smaller than the first, so their effect is less important. Market is included.

Applying the same procedure to the data without the variable corresponding to the market induces the same results and nearly the same values in the tail dependence matrix, as seen in Figure 4.5 and Figure 4.6. We keep the market to show which sectors are the most dependent on the whole market index.

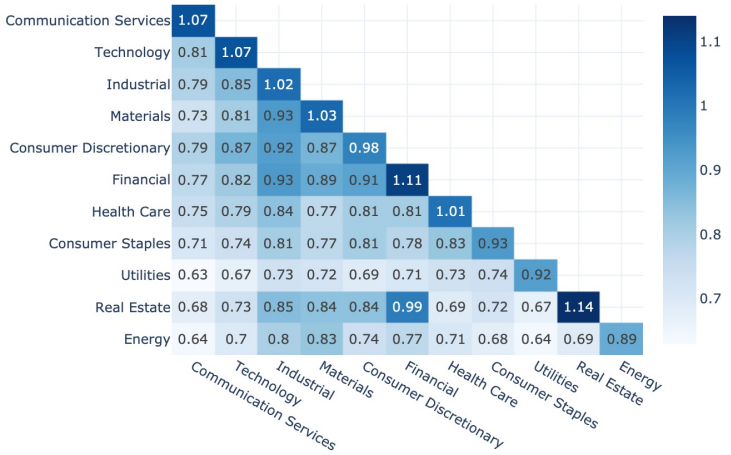

Figure 4.5: Empirical estimate of the tail pairwise dependence matrix based on the polar transformation of losses and using the observations with the largest 5% of Euclidean norms. Market is not included.

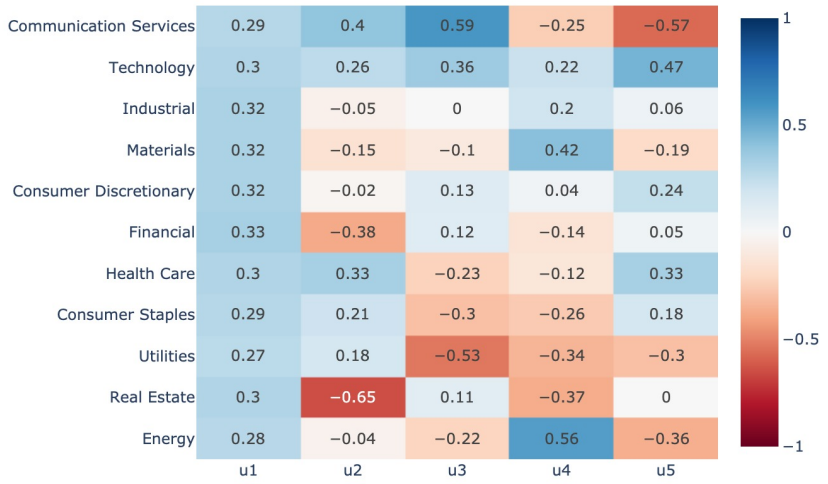

Figure 4.6: First five eigen-vectors, i.e. loadings, corresponding to the eigen-decomposition of the tail dependence matrix $\hat{\Sigma}_Z$. Market is not included.

This approach of summarizing extremal dependence is quite interesting since it resembles the covariance matrix. It also provides a simple visualization tool for high dimensional multivariate analysis. Since the tail dependence matrix is positive semi-definite, it may be adapted to be used in portfolio optimization instead of the covariance matrix, for example.

4.2 Graphical Lasso

We use the tail dependence matrix to compute the partial correlation matrix. If the angular components had a joint Gaussian distribution, then the partial correlation can be extracted from the tail dependence matrix, and a graphical model describes the dependence structure of the sectors. Based on this assumption and since the tail dependence matrix is positive semi-definite, if $K = \hat{\Sigma}_Z^{-1}$ is the precision matrix, then the partial correlation matrix is

$$\rho = -\text{diag}(K)^{-1/2} \times K \times \text{diag}(K)^{-1/2},$$

and $\rho_{ii} = 1$ for $i = 1, \ldots, d$ by definition.

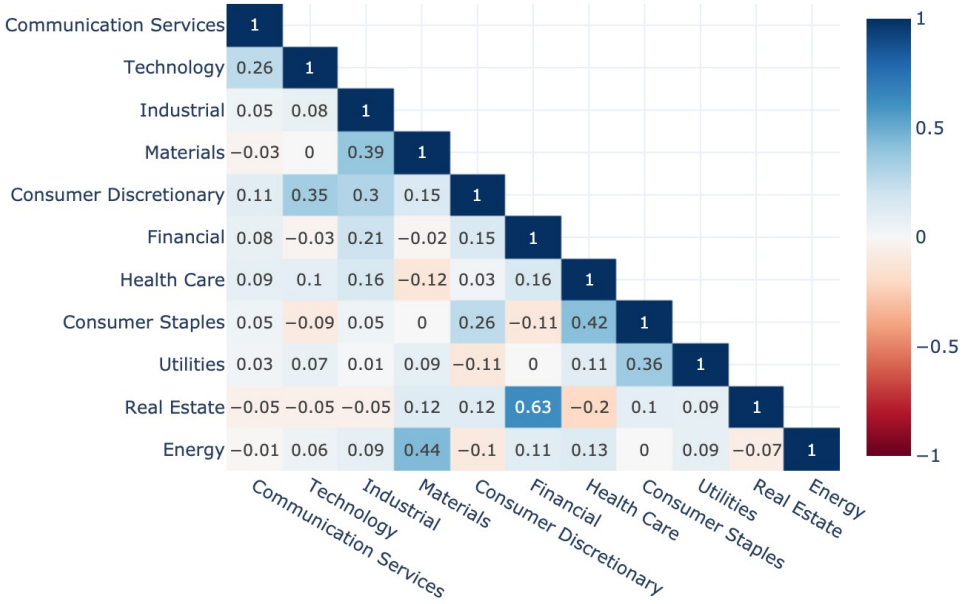

Figure 4.7: Partial correlations based on the inversion of the tail dependence matrix. Large values in absolute norm imply strong dependence between the corresponding pairs of sectors conditioning on all other variables.

Figure 4.7 shows the estimated partial correlation matrix based on the tail dependence matrix in Figure 4.5. Small negative values may indicate non-significant dependence but the negative value of -0.2 between Real Estate and Health Care does not make sense and may be due to inappropriate assumption of normality. Otherwise, this means there is an inefficiency, and possibly an arbitrage opportunity, in the market. We are not happy with the results because of the negative values and the lack of sparsity in the matrix. We could use an arbitrary threshold to make the matrix sparse but the selection of such a threshold is problematic. We decided to experiment with the **graphical lasso**, i.e. glasso, algorithm (Friedman, Hastie, and Tibshirani 2008), which is implemented in **scikit-learn** (Pedregosa et al. 2011). By introducing a penalization parameter α, the algorithm introduces sparsity in the covariance matrix.

In our case, we apply glasso on the tail dependence matrix $\hat{\Sigma}_Z$, then compute the partial correlation as before. To select α, we first used cross-validation based on the likelihood score but it didn't perform well at all. The second approach is to select a sufficiently high α in order to have sparsity without losing too much information. Figure 4.8 describes the change in sparsity depending on the selected penalization. The regularization starts to introduce sparsity from $\alpha = 0.5$. For $\alpha = 0.8$, we see that more than half of the partial correlations are zero. Figure 4.9 shows the corresponding partial correlation, estimated using glasso for $\alpha = 0.8$. The strongest dependence is between Real Estate and Financial, which is expected based on the tail dependence matrix. Utilities have no dependence on the other sectors. Communication Services is dependent only with Technology. Energy is also only dependent on Materials. Both pairs showcase a partial correlation close to zero.

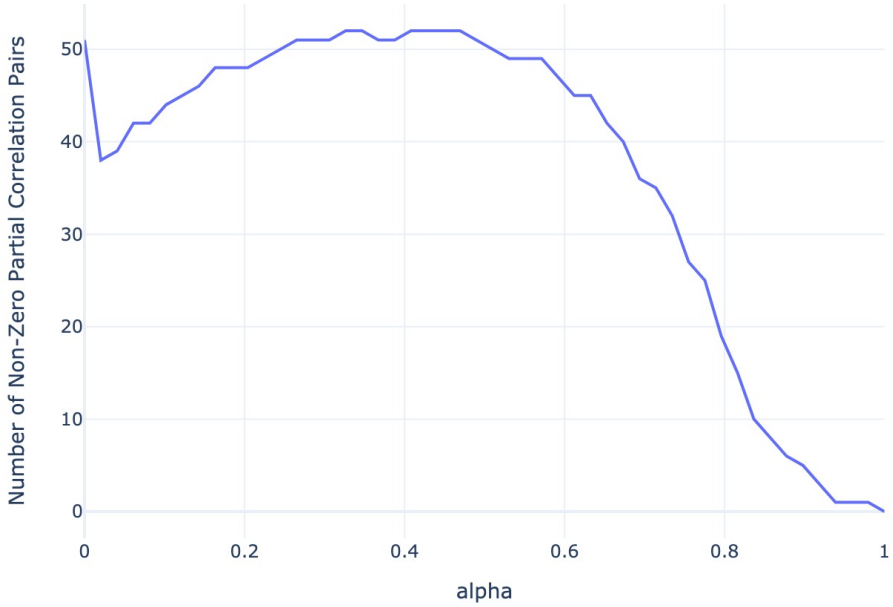

Figure 4.8: Number of non-zero pairs in the partial correlation matrix against the regularization parameter α, which should be high enough to introduce sparsity and simplify the dependence structure of the sectors.

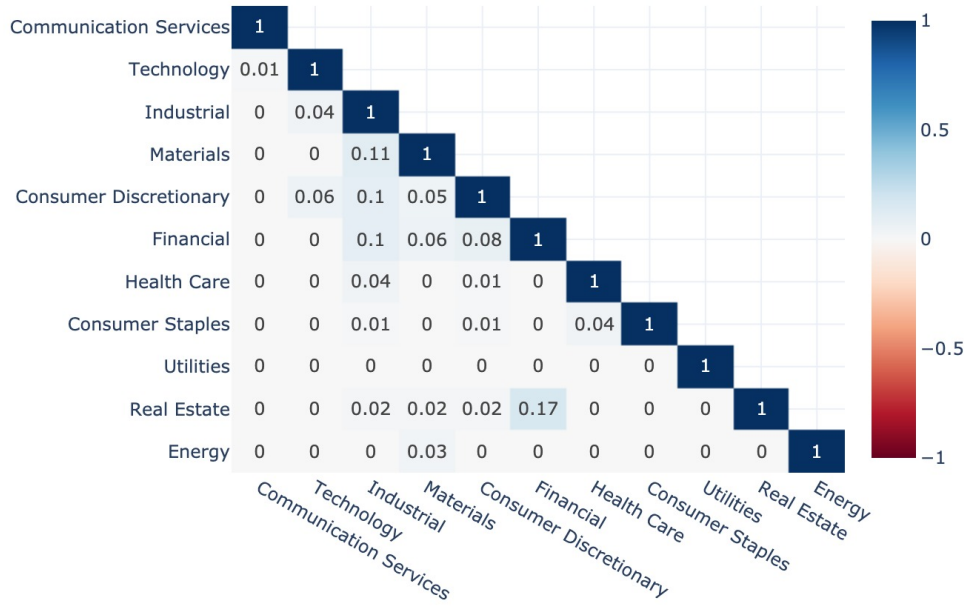

Figure 4.9: Partial correlations estimated from the graphical lasso estimation of the sparse tail dependence matrix, where $\alpha = 0.8$.

Figure 4.10 represents the graphical model of the dependence structure between the sectors for two values of α. It is based on the partial correlation, which is computed by introducing sparsity into the tail dependence matrix using the graphical lasso and inverting the estimated matrix. An edge between two sectors represents a non-zero partial correlation between them. We kept the same color scheme as in Figure 3.9 but it shouldn't be considered as actual clusters that can be derived from partial correlation. The four sectors in red are still dependent even for $\alpha = 0.85$ and share a connection with Technology and Real Estate. To conclude, the same dependence structure keeps appearing even if the dependence measure or approach changes. Is it enough to say that we discovered the real subgroups of the S&P 500 or are we missing something and keep convincing ourselves of these results?

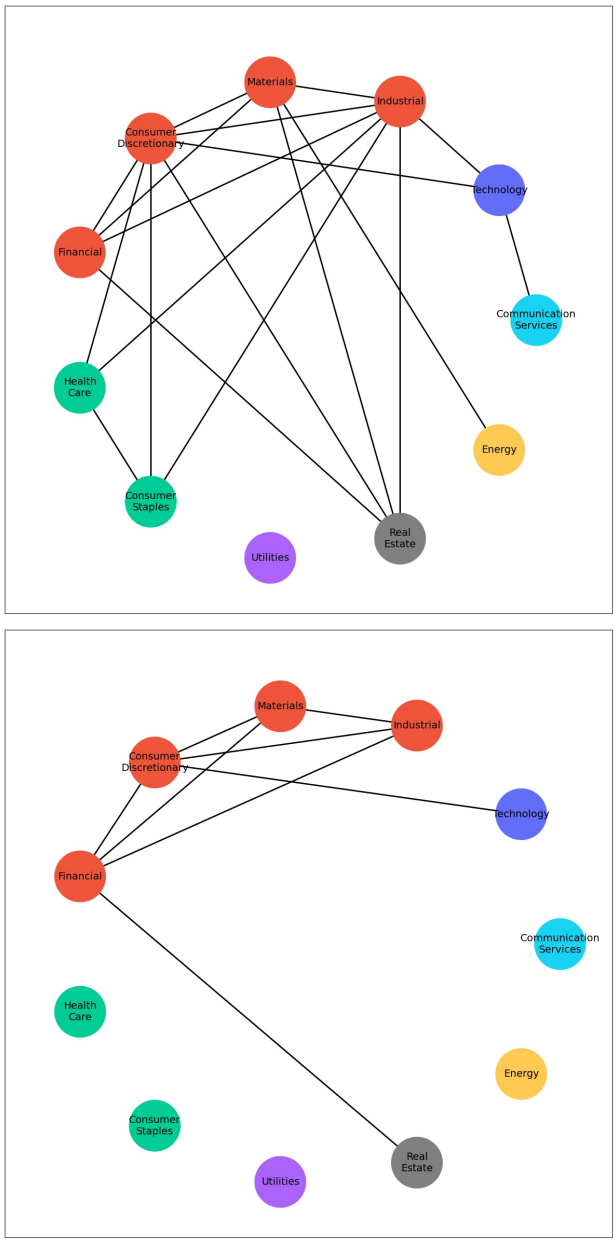

Figure 4.10: Plot of the sector network based on the estimated partial correlation. Sparsity derives from graphical lasso for $\alpha = 0.8$ (top) and $\alpha = 0.85$ (bottom). Nodes represent sectors. Edges correspond to non zero partial correlation. Colors are the same as in Figure 3.9.

4.3 Clustering of Extremes

Based on the previous analyses, we observed that some sectors behave similarly in terms of extreme losses and are dependent. We talked about possible subgroups or clusters of sectors in Figure 3.9 and in Figure 4.1. In this section, we demonstrate how the spherical k-means, a variant of k-means clustering algorithm, can be applied in the analysis of extreme observations from the data, as described in Janßen and Wan (2020). The technique is used to find clusters of extremal dependent sectors, detect relevant patterns during extreme losses and classify them.

Using a clustering algorithm on the observations having the largest norm is naive since it is inefficient, because the extremal points tend to be spread out and if the observations are heavy-tailed then problems arise from the possibly infinite second moments. However, by providing a theoretical background on how to use the spherical k-means, the algorithm estimates concrete measures related to extremal events. First, we briefly present a spherical k-means algorithm. k-means clustering aims to partition observations into k clusters, with each observation belonging to the cluster with the nearest mean, called cluster centers or centroids, which serve as a prototype of the cluster. The usual distance used between observations is Euclidean, but there are alternatives. Since the focus is on extremal dependence, in particular the spectral measure, and we use a sample, i.e. w, from a quadrant of the unit sphere $\mathbb{S}_+^{d-1} = \{x \in [0,\infty)^d : ||x|| = 1\}$, the angle between two points is a more appropriate distance. The spherical k-means procedure is a modification of the original k-means, which measures differences in terms of angular dissimilarity

$$d(x,y) = 1 - \cos(x,y) = 1 - \frac{\langle x,y\rangle}{||x||_2||y||_2}.$$

In this case, the estimated cluster centers are on $\mathbb{S}_+^{d-1}$ and are only determined by their direction. We use the package by Hornik et al. (2012) for the estimations.

Now, we introduce the estimation procedure. We standardize the margins by transforming them to standard Fréchet(1) using the transformation

$$Z_i = \frac{1}{1 - F_i(X_i)},$$

where F_i are empirically estimated. We compute the radial and angular components

$$R_i = ||Z_i||_2 \quad \text{and} \quad W_i = \frac{Z_i}{R_i},$$

and select a sufficiently high threshold for r, denoted as r_0. The component w represents the projection of the transformed observations onto the unit sphere $\mathbb{S}_+^{d-1}$. The spherical k-means clustering algorithm is applied to the observations $w_1, \ldots, w_n$ of W for which $r_i > r_0$. For a specified k, we obtain k cluster centers that can be interpreted as a dependence prototype for a particular class of an extremal event.

We apply the above framework to the daily losses. We choose the threshold r_0 corresponding to the 95% quantile of r. The second choice we have to make is the number of clusters k but it is rather difficult to find a concrete value. We apply the method for $k = 4, \ldots, 7$ and compare the results with the exploratory analysis, mainly the dependence structure in Figure 3.9 and 4.1. We also keep in mind the macro-economic events of the previous twenty years. We summarize the results for the different k values.

- $k = 4$: In Figure 4.11, the first cluster C_1 signals the asymptotic independence of Utilities and Consumer Staples from the other sectors. C_2 does the same thing for the Real Estate and Financial sectors and C_4 for Communication Services and Technology. The last cluster C_3 contains the rest.

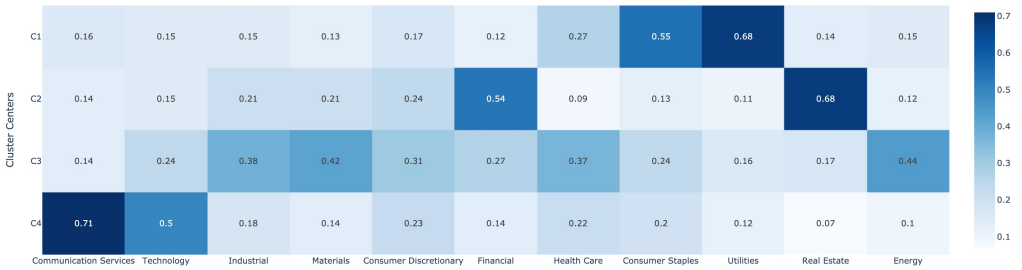

Figure 4.11: Estimated spherical k-means cluster centers on the S&P 500 losses data for $k = 4$. Cluster centers are on the unit sphere.

- $k = 5$: In Figure 4.12, the Real Estate and Financial cluster is still present, as C_2. Communication Services and Utilities are in separate clusters, respectively C_3 and C_5, and asymptotically independent from each other and other sectors. C_1 contains Materials and Energy. C_4 encompasses the rest. We observe that Technology's dependence structure is now different.

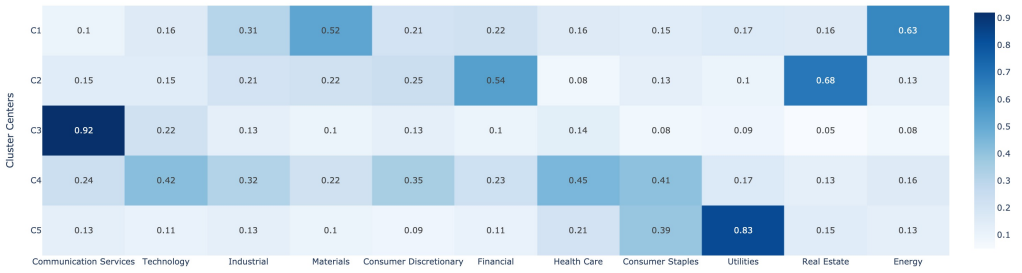

Figure 4.12: Estimated spherical k-means cluster centers on the S&P 500 losses data for $k = 5$. Cluster centers are on the unit sphere.

- $k = 6$: In Figure 4.13, we observe the same structure except for Health Care and Consumer Staples, which are now in the same cluster. Technology behaves differently once again.

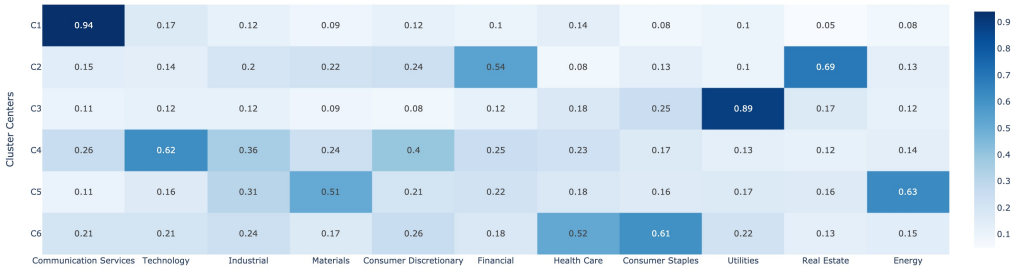

Figure 4.13: Estimated spherical k-means cluster centers on the S&P 500 losses data for $k = 6$. Cluster centers are on the unit sphere.

- $k = 7$: This is the first number of clusters that we can defend based on the exploratory analysis. We got nearly the same clusters as in Figure 3.9. In Figure 4.14, we have Utilities in C_1, Communication Services in C_2, Energy in C_3 and Technology in C_5. C_6 contains Real Estate and Financial, which is more obvious than what the pairwise dependence measures suggested in exploratory analysis. C_7 contains Health Care and Consumer Staples. The last cluster C_4 encompasses the rest, in particular the most dependent sectors and the most dependent on the whole market.

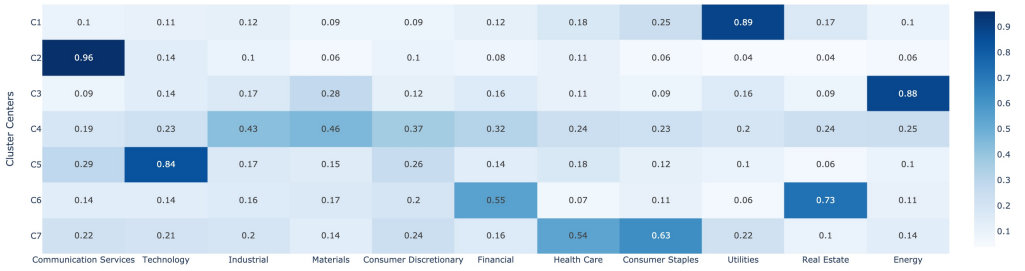

Figure 4.14: Estimated spherical k-means cluster centers on the S&P 500 losses data for $k = 7$. Cluster centers are on the unit sphere.

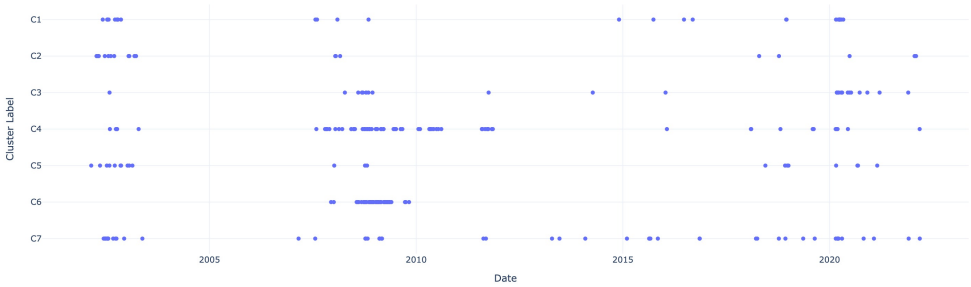

Figure 4.15: Estimated spherical k-means cluster labels time series on the S&P 500 losses data for $k = 7$. The y-axis represents the classification label, and the x-axis corresponds to the date. There is a unique point for each date that is placed on the horizontal line corresponding to its label. Successive points highlight macroeconomic events.

We focus on the case $k = 7$. As a remark, we applied the same procedure but included the market as a variable. The same cluster centers are obtained and the market is most probably in the cluster previously called C_4, corresponding to C_3 in Figure 4.16.

Figure 4.16: Estimated spherical k-means cluster centers on the S&P 500 losses data for $k = 7$ including the market. Cluster centers are on the unit sphere.

Each daily observation is labeled, detecting which sectors suffered the most from extreme losses. Figure 4.15 shows the timeline plot of cluster labels. There is no overlapping between the points, with each observation plotted on the horizontal line corresponding to its cluster and a vertical line corresponding to its date. This representation allows us to identify major events in financial market history. We notice the effect of the dot-com bubble in 2002 and 2003, where only sectors related to innovation and technology experienced major losses. The underlying stocks were trading at high price to earnings ratios and with high growth expectations. We add to that the corporate fraud scandals of 2002, such as Enron, along with the loss of confidence from investors due to 9/11 and the Iraq war. Starting from 2008, the Real Estate and Financial sectors (cluster C_6) experienced successive days of losses. The cluster C_4, corresponding to the sectors that are the most dependent on the whole market, experienced much larger losses than the other clusters. In the case of the COVID crisis, we notice that the Real Estate and Financial cluster C_6 did not have any observations. This doesn't necessarily mean the sectors didn't experience huge losses but may imply the other sectors had larger losses on the same days.

4.4 Conditioning on the Market

In this section we present a conditional multivariate extreme value model that we find interesting. We saw that the sectors are significantly dependent on the whole market but the strength of dependence varies. We want to quantify the behavior of the sectors conditioning on the market having high enough losses. We used the framework developed by Heffernan and Tawn (2004) implemented in the R package **texmex** (Southworth, Heffernan, and Metcalfe 2020). The method works on high-dimensional data and takes care of the observations where not all variables are extreme at the same time, which we discussed previously in the asymmetry section and saw in Figure 3.4. For example, Utilities shows large losses even if the market didn't. and inversely.

We briefly present how the model is constructed. First, GPD models are fitted to the marginal variables, and then are transformed to Laplace margins. The GPD is used to transform the observations above a selected threshold and the observations below the threshold are transformed empirically by a rank transform, equivalent to what is done in Figure 3.5, except it transforms the margins to the Laplace scale instead of the Fréchet. We denote the transformed margins $Y_i = G^{-1}(\hat{F}_i(X_i))$, for $i = 1, \ldots, d$, where G denotes the Laplace distribution function. The original version of the model uses Gumbel margins instead of Laplace. The theoretical motivation is presented in Heffernan and Tawn (2004) and is mainly based on the limit representation and its properties. The paper by Keef, Papastathopoulos, and Tawn (2013) presents the version using Laplace margins; in particular the authors describe what is equivalent and different from using Gumbel margins. One of the main advantages is that the symmetry of the Laplace distribution ensures the limiting dependence model is unchanged simply by inverting the copula to produce a model for negatively associated variables. Let Y_m be the marginal variable corresponding to the market. Conditional on Y_m exceeding a high threshold u_m, the remaining vector Y_{-m}, i.e. the sectors, takes the form

$$Y_{-m} = \alpha Y_m + Y_m^{\beta} Z,$$

where $Z \in \mathbb{R}^{d-1}$ is the residual vector, and the vector $\alpha \in [-1, 1]^{d-1}$ and $\beta \in (-\infty, 1)^{d-1}$ contain the parameters.

The choice of threshold for the GPD transform is based on our exploratory analysis, and we use $u = 2$ for all sectors. In the case of the dependence model, where we have to select a threshold for the market, we used the diagnostic tools available in the package. The model is fitted for a range of thresholds, and then the estimated parameters are plotted against the corresponding quantile. Above a sufficiently high threshold, the parameters should remain constant. The optimization procedure, i.e. likelihood maximization, is sensitive to the initial values of the parameters, so we had to select appropriate starting values to stabilize convergence. Figure 4.17 to 4.20 indicate that for most sectors the model fails to fit the data. For α, we saw an increase in the estimate with the increase in the conditioning threshold for most sectors, except the very dependent to the market (Industrial), where the parameter remains constant. This is coherent since Figure 3.5 shows how the observations are closer to the diagonal when looking to the extremes, which implies a linear relation between the Industrial sector and the market. There isn't an appropriate and sufficiently high threshold that stabilizes the parameter estimates, in particular β. One possible explanation is the change in dispersion when changing the threshold, which was seen in Figure 3.5; β captures this dispersion in some sense. However, the profile likelihood plots show that the estimation actually corresponds to the maximum of the objective function.

After fitting the model, we use the diagnostic plots provided for validation. For each sector, we have three plots: The first graphs the centered and scaled values of the dependence model residuals against the extreme conditioning variable, and the second is the centered absolute value of the first. If the models fit the data, these two plots shouldn't show structure and the smoothers should be horizontal. The third shows the original data with quantiles of the fitted conditional model, which should agree with the shape of the data. Figure 4.21 to 4.24 show that the two residual plots are acceptable for most sectors, except Real Estate. However, the quantiles seem a bit different from the shape of the original data, especially close to the market threshold when sectors have larger losses.

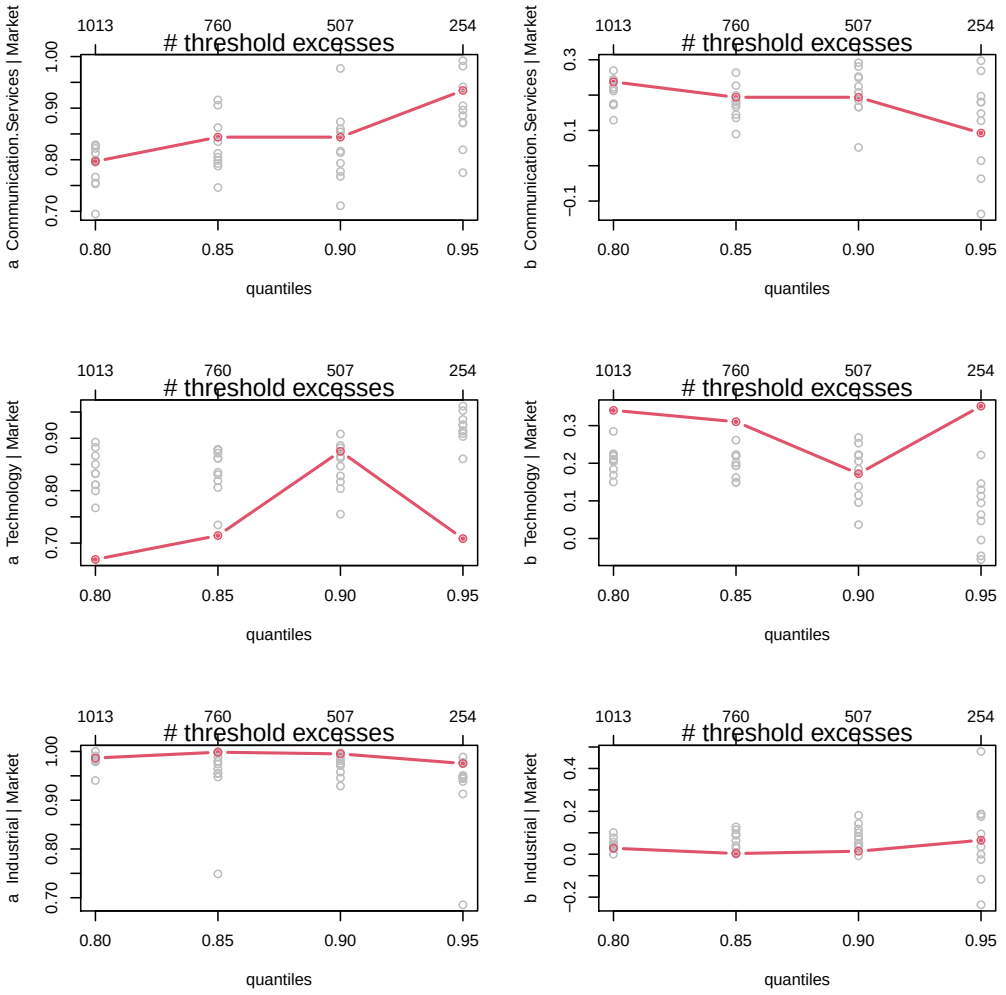

Figure 4.17: Part 1. Market threshold selection diagnostics for the dependence model. The y-axis represents the four thresholds that are tested, equal to the 0.8, 0.85, 0.9 and 0.95 quantiles. The red points correspond to the model parameter estimates, and the grey points are the estimates obtained via bootstrapping, and the values are on the y-axis. The rows represent sectors and columns the parameters α and β. Stability of the parameters indicates a good fit.

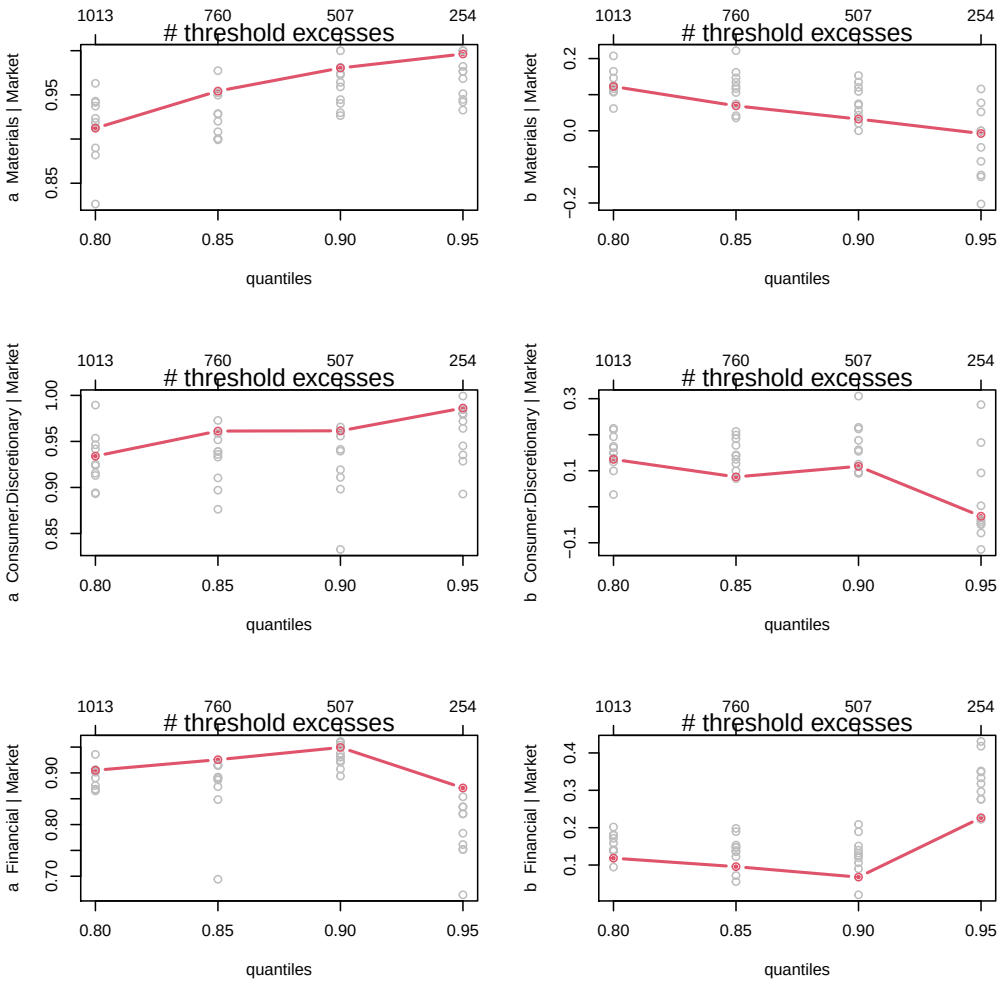

Figure 4.18: Part 2. Market threshold selection diagnostics for the dependence model. The y-axis represents the four thresholds that are tested, equal to the 0.8, 0.85, 0.9 and 0.95 quantiles. The red points correspond to the model parameter estimates, and the grey points are the estimates obtained via bootstrapping, and the values are on the y-axis. The rows represent sectors and columns the parameters α and β. Stability of the parameters indicates a good fit.

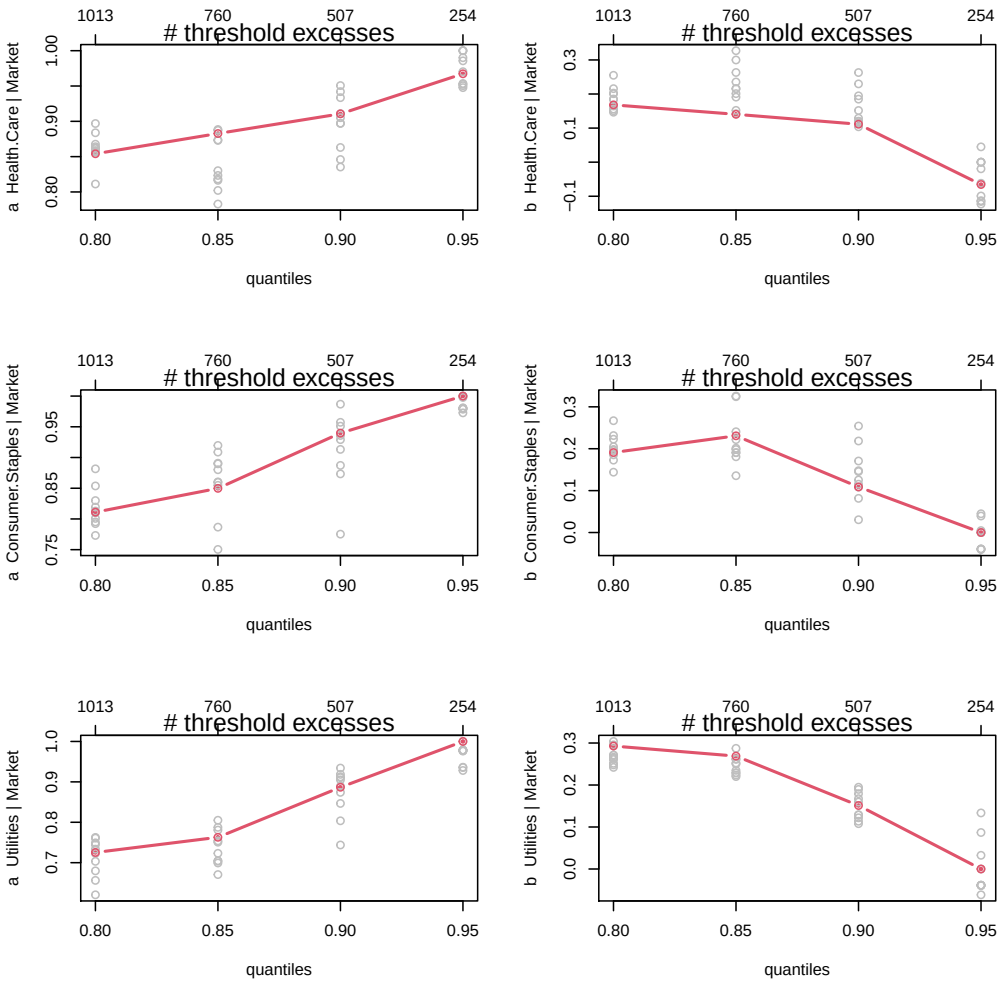

Figure 4.19: Part 3. Market threshold selection diagnostics for the dependence model. The x-axis represents the four thresholds that are tested, equal to the 0.8, 0.85, 0.9 and 0.95 quantiles. The red points correspond to the model parameter estimates, and the grey points are the estimates obtained via bootstrapping, and the values are on the y-axis. The rows represent sectors and columns the parameters α and β. Stability of the parameters indicates a good fit.

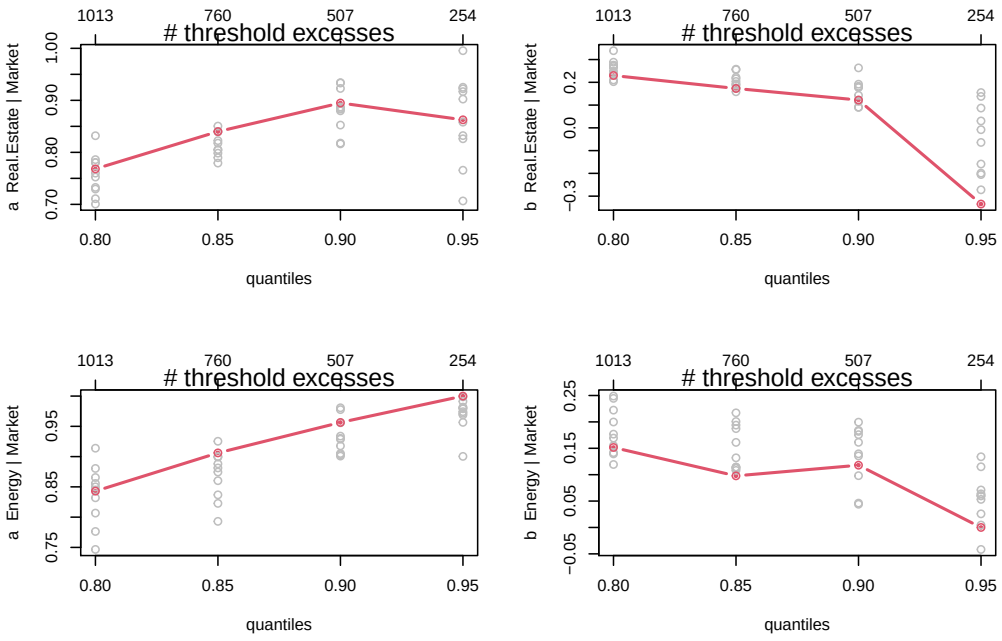

Figure 4.20: Part 4. Market threshold selection diagnostics for the dependence model. The y-axis represents the four thresholds that are tested, equal to the 0.8, 0.85, 0.9 and 0.95 quantiles. The red points correspond to the model parameter estimates, and the grey points are the estimates obtained via bootstrapping, and the values are on the y-axis. The rows represent sectors and columns the parameters α and β. Stability of the parameters indicates a good fit.

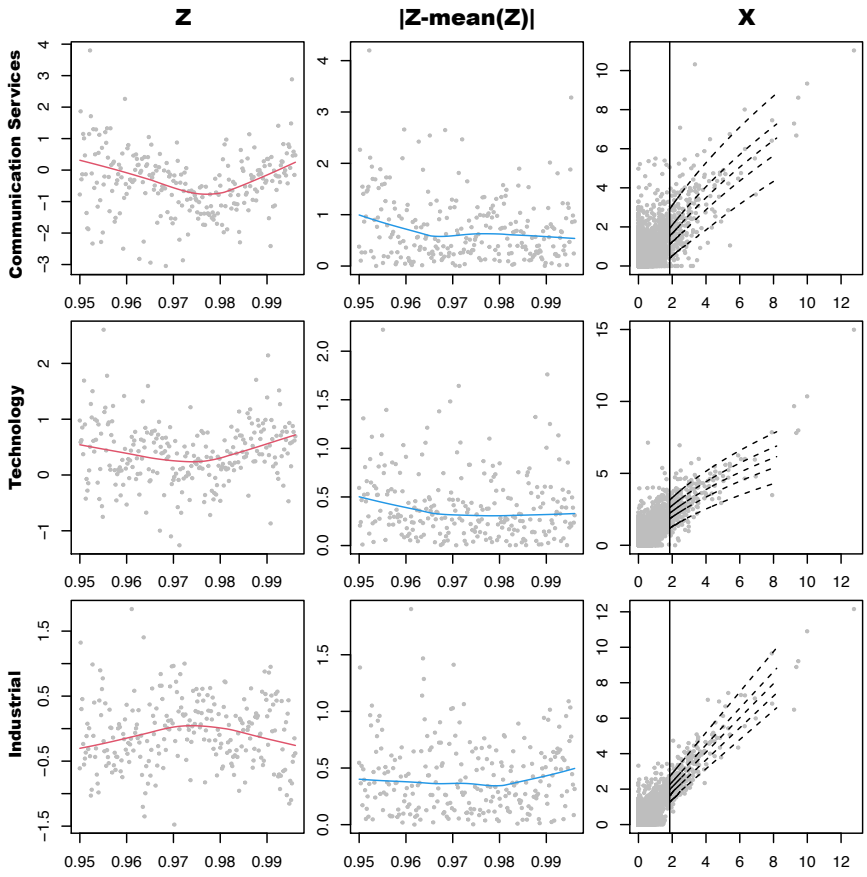

Figure 4.21: Part 1. Conditional model diagnostic, where $q_m = 0.95$. Left figure is the plot of centered and scaled values of the dependence model residuals, i.e. Z, against the conditioning variable, i.e. Y_m. Center figure is the centered absolute values of the first, i.e. $|Z - \text{mean}(Z)|$. Right figure is the plot of original data with quantiles, respectively $0.1, 0.3, 0.5, 0.7$ and 0.9, of the fitted conditional model. Row order: Communication Services, Technology and Industrial.

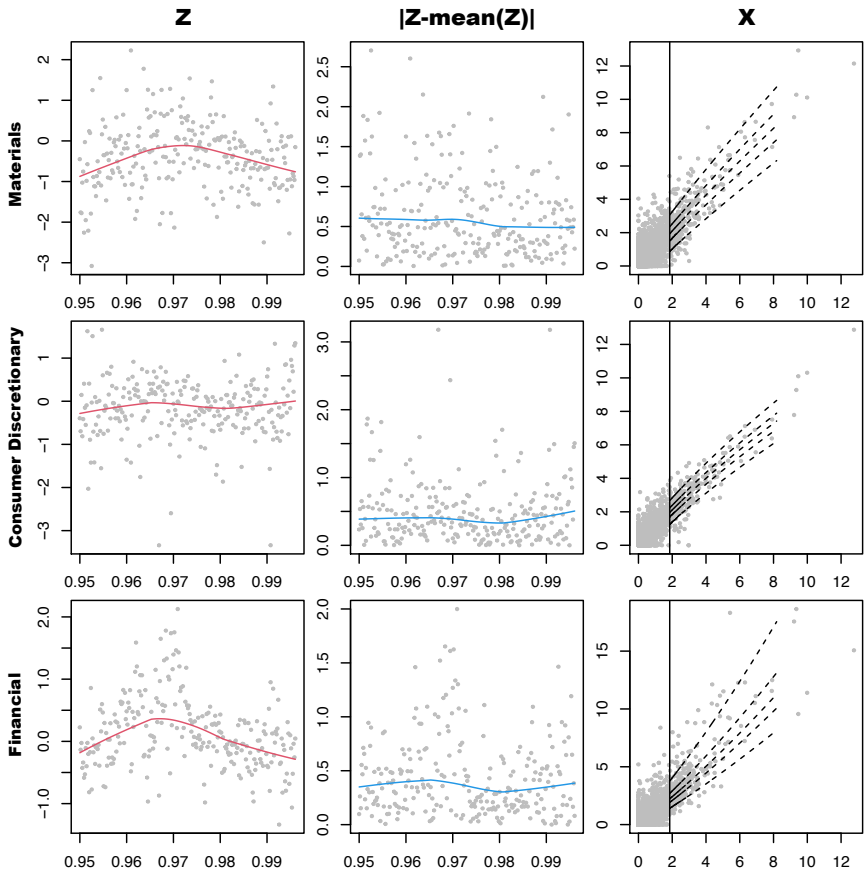

Figure 4.22: Part 2. Conditional model diagnostic, where $q_m = 0.95$. Left figure is the plot of centered and scaled values of the dependence model residuals, i.e. Z, against the conditioning variable, i.e. Y_m. Center figure is the centered absolute values of the first, i.e. $|Z - \text{mean}(Z)|$. Right figure is the plot of original data with quantiles, respectively $0.1, 0.3, 0.5, 0.7$ and 0.9, of the fitted conditional model. Row order: Materials, Consumer Discretionary and Financial.

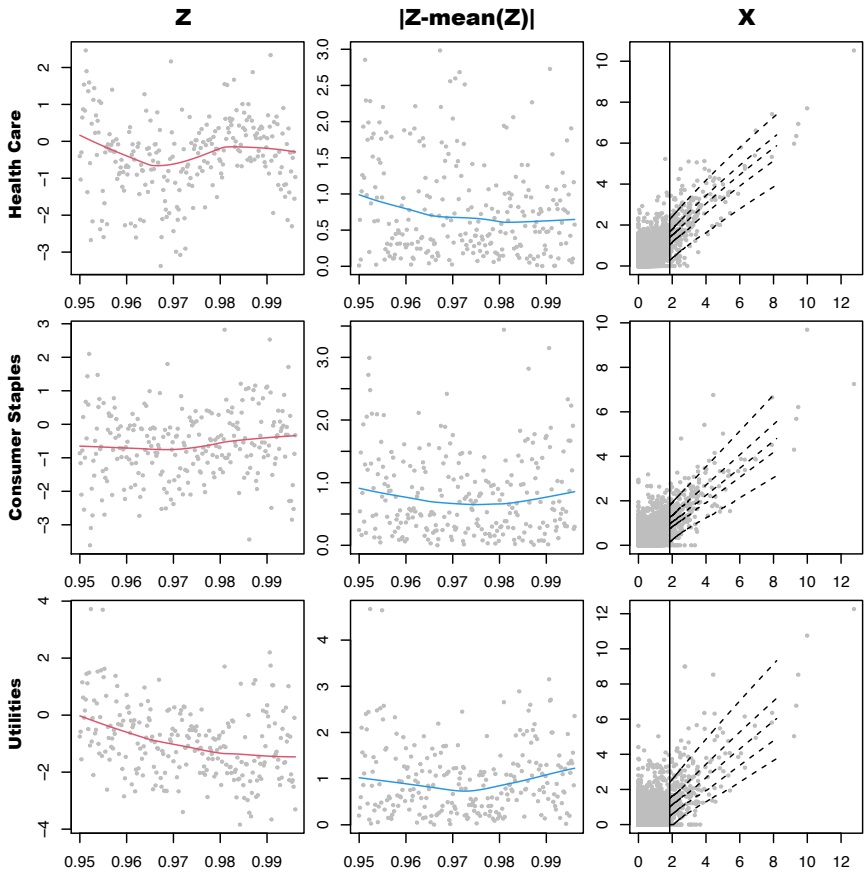

Figure 4.23: Part 3. Conditional model diagnostic, where $q_m = 0.95$. Left figure is the plot of centered and scaled values of the dependence model residuals, i.e. Z, against the conditioning variable, i.e. Y_m. Center figure is the centered absolute values of the first, i.e. $|Z - \text{mean}(Z)|$. Right figure is the plot of original data with quantiles, respectively $0.1, 0.3, 0.5, 0.7$ and 0.9, of the fitted conditional model. Row order: Health Care, Consumer Staples and Utilities.

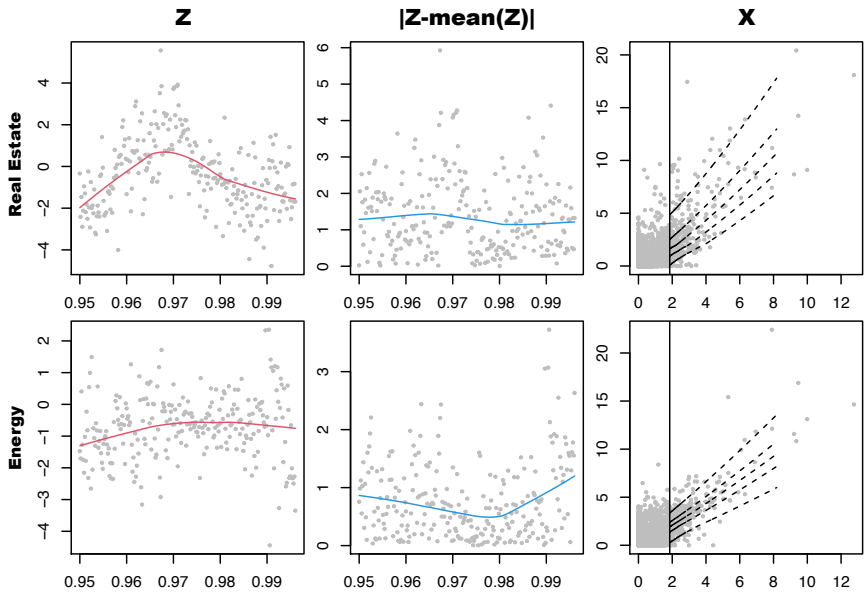

Figure 4.24: Part 4. Conditional model diagnostic, where $q_m = 0.95$. Left figure is the plot of centered and scaled values of the dependence model residuals, i.e. Z, against the conditioning variable, i.e. Y_m. Center figure is the centered absolute values of the first, i.e. $|Z - \text{mean}(Z)|$. Right figure is the plot of original data with quantiles, respectively $0.1, 0.3, 0.5, 0.7$ and 0.9, of the fitted conditional model. Row order: Real Estate and Energy.

Figure 4.25 shows the parameter estimates for different dependence model thresholds. Clearly there is a big difference, in particular for β, where some values become negative when increasing the threshold. This is clearly a sign that the model is not appropriate for this dataset. Even the relative position of sectors changes, thus no clear dependence structure with the market can be extracted. One positive aspect of the model is that ordering the sectors by α is coherent with the other dependence measures.

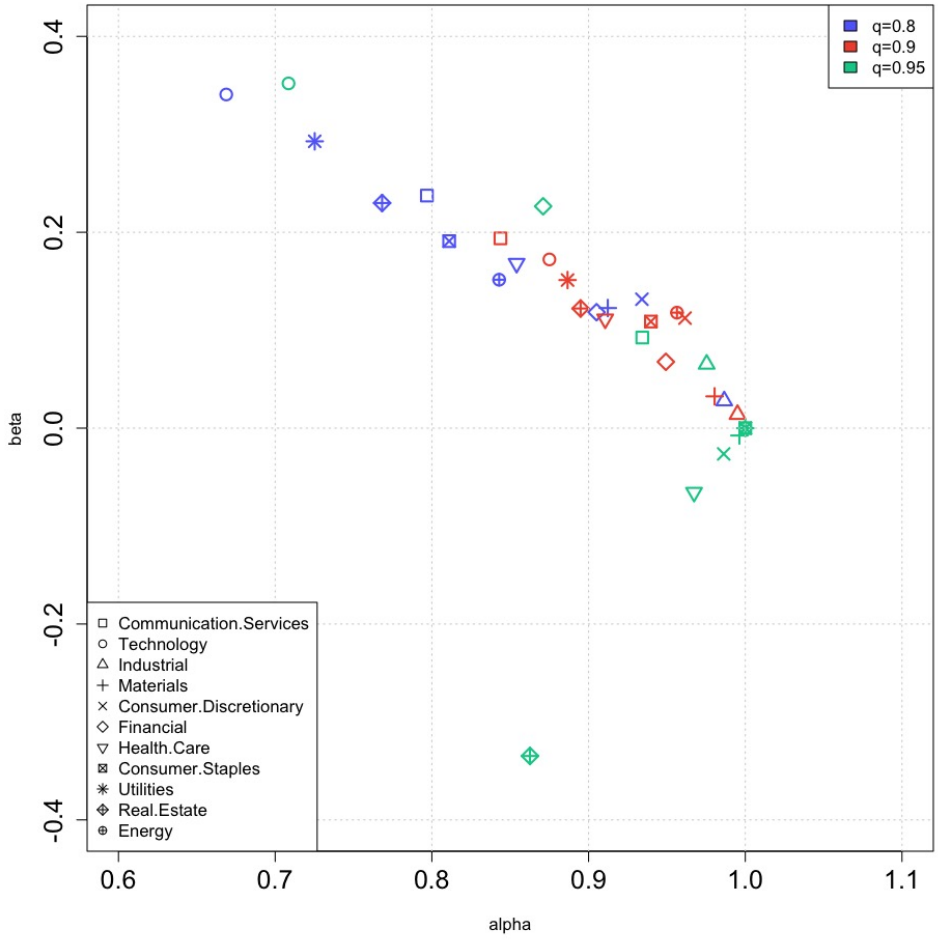

Figure 4.25: Conditional model parameters estimate. Three quantiles for conditioning variable threshold.

To conclude this section, the idea of conditioning on the market is interesting and easy to interpret but the tested model performed poorly. The α estimates, indicating the strength of dependence of the sectors to the market, were close to previous analyses. However, the model failed in describing the dispersion in the extremes, based on the parameter β. The assumption of normality in the residuals may be not appropriate or the linearity of the model may be too simple. However, this simplicity is the main advantage of the model since it implies easier interpretation and fast computation.

Summary

In terms of dependence between the S&P 500 and its sectors, the three approaches shared a similar structure. Industrial, Materials and Consumer Discretionary are the most dependent on the market. Utilities and Communication Services are the least dependent on the market. This was also the case in the bivariate analysis, where we looked at several dependence measures. Exploratory analysis, in particular in the extremal dependence section, agrees with the multivariate analysis of the sectors. The structure in Figure 3.9 and Figure 4.10 is similar, especially the red cluster. The main difference is the very strong dependence between Real Estate and Financial, which wasn't that obvious before doing a multivariate analysis.

5. Conclusion

When starting this project, we didn't have in mind a specific objective or model to test on the S&P 500 sector indices dataset. We focused on a historical analysis of these sectors, which provides both qualitative and quantitative insights. Since we were interested in large losses, that rarely happen for such diversified indices, we used extreme value theory to build such insights. It was important for us to have a solid mathematical background and clear assumptions for this analysis, so we made sure the report is self-contained and explains the theory behind every model, indicator and figure. Also, the quality of the figures and their ease of interpretation were a priority. The scripts constructing the figures are available on GitHub. Most of the papers used as references present a theoretical framework, then apply it on two or three different datasets. Inversely, this report compares different approaches of extremes analysis on the same dataset.

The first chapter presented the reasons behind the choice of the dataset. Briefly, we believe balancing between the S&P 500 sectors is an interesting portfolio for investors with higher risk appetite but won't do the effort of selecting single stocks one by one. Reading balance sheets and predicting future revenues is no longer necessary for these investors. In chapter two, we presented a theoretical background related to extreme value theory.

In chapter three, we tackled exploratory data analysis. We started with the univariate analysis, i.e. the margins, where we decide above which threshold losses are considered as extreme. It was challenging since the quantiles of each sector were different, but in the end we decided that a loss of 2% or more is an appropriate threshold, which is close to the 95% quantile for most sectors. After that, we focused on the bivariate analysis, i.e. all pairs of sectors, which gives a general idea on what the multivariate extremal de-

pendence structure might look like. At that time, we decided that describing the extremal dependence of the S&P 500 sectors should be the priority of this project. Several dependence measures were used for this purpose such as the extremal correlation χ, the coefficient of tail dependence η and Pickands dependence function A. Even though the approaches were different, the measures shared the same results. We noticed a certain asymmetry between the sectors, so we tried to quantify that using the coefficient of extremal asymmetry, as defined by Semadeni (2020).

The fourth and last chapter focused on multivariate analysis. It is divided into three sections. The first is a framework, described in Cooley and Thibaud (2019), that defines a high-dimensional tail dependence matrix, which is positive semi-definite and can be seen as a covariance matrix for extremes. An eigen-decomposition of the tail dependence provides tools for exploring extremal dependence, which looks like an adaptation of principal component analysis to extreme value theory. We used the matrix to create a graphical model that describes the dependence structure of the sectors. We also used graphical lasso to introduce sparsity in the graph by keeping the most significant dependencies. This allowed us to extract a simple network describing the dependence structure of the sectors. In the second section, we used the spherical k-means algorithm to identify subgroups or clusters of sectors based on extremal dependence, as described in Hornik et al. (2012). It allows to classify losses and detect important financial events, such as the 2008 financial crisis. These two sections were mainly based on the angular measure and the polar representation of the data, i.e. in terms of norm and angle. The last section was a bit different. Since the dependence between the sectors and the whole market varies, we tested a multivariate conditional model by Heffernan and Tawn (2004), where the conditioning variable is the S&P 500 index. The model is designed to take care of the extremal dependence. Unfortunately, we got poor results from the model.

The dependence structure of the S&P 500 market can be summarized as follows. The most dependent sectors to the market are Industrial, Consumer Discretionary and Financial. The least dependent sectors to the market are Communication Services and Utilities. The most dependent pairs of sectors are Industrial with Materials, Consumer Discretionary and Financial based on all measures. We add to that the Financial/Real Estate and Energy/Materials pairs, based on the multivariate analysis. The least dependent

pairs of sectors are Communication Services and Utilities with nearly all the other sectors. It is also the case for the pair Real Estate and Energy. The cluster centers obtained from the spherical k-means procedure agree with these results. It also confirms a dependence between Health Care and Consumer Staples.

We observed during the first semester of 2022 some interesting patterns. We saw many trading days where oil prices went up and at the same time the technology sector suffered losses. Also, days where nearly all the sectors went down and the next day financial stocks gained much more than the other stocks. This means that the dependence could be lagged in time. In this research, we didn't specifically approach the data as a time series. We transformed the indices time series to have daily losses and considered them as independent observations. It should be interesting to see the effect of time on the dependence structure. First, by simply looking at the variations in the structure in time. Second, we could look at the dependence not for daily losses but a longer time frame. We already tried this when estimating the Pickands function for the maxima over a week.

Finally, I wish to express my sincere appreciation to my supervisors, Dr. Anthony C. Davison and Mario Krali, who gave me the opportunity to undertake this master's thesis within the chair of statistics at EPFL. My gratitude goes to them for allowing me to learn so much about extreme value theory, to answer all my questions and for all their suggestions and recommendations throughout the project. I would also like to thank everyone involved in my studies at EPFL, who contributed over the years making the work environment stimulating.

Bibliography

Coles, Stuart (2013). *An Introduction to Statistical Modeling of Extreme Values*. Springer London, Limited.

Engelke, Sebastian and Jevgenijs Ivanovs (2021). "Sparse Structures for Multivariate Extremes". In: *Annual Review of Statistics and Its Application* 8.1, pp. 241–270. DOI: 10.1146/annurev-statistics-040620-041554. eprint: https://doi.org/10.1146/annurev-statistics-040620-041554. URL: https://doi.org/10.1146/annurev-statistics-040620-041554.

Hornik, Kurt et al. (2012). "Spherical k-Means Clustering". In: *Journal of Statistical Software* 50.10, pp. 1–22. DOI: 10.18637/jss.v050.i10.

Jolliffe, I. T. (2002). *Principal component analysis.* English. 2nd ed. Springer Ser. Stat. New York, NY: Springer. ISBN: 0-387-95442-2. DOI: 10.1007/b98835.

Wang, Jiguang (2013). "Partial Correlation Coefficient". In: *Encyclopedia of Systems Biology*. Ed. by Werner Dubitzky et al. New York, NY: Springer New York, pp. 1634–1635. ISBN: 978-1-4419-9863-7. DOI: 10.1007/978-1-4419-9863-7_373. URL: https://doi.org/10.1007/978-1-4419-9863-7_373.

Banerjee, Onureena, Laurent El Ghaoui, and Alexandre d'Aspremont (2008). "Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data". English. In: *J. Mach. Learn. Res.* 9, pp. 485–516. ISSN: 1532-4435. URL: www.jmlr.org/papers/v9/banerjee08a.html.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani (2008). "Sparse inverse covariance estimation with the graphical lasso". English. In: *Biostatistics* 9.3, pp. 432–441. ISSN: 1465-4644. DOI: 10.1093/biostatistics/kxm045.

Heffernan, Janet E. and Alec G. Stephenson (2018). "ismev: An Introduction to Statistical Modeling of Extreme Values". In: R package version 1.42. URL: https://CRAN.R-project.org/package=ismev.

Stephenson, A. G. (2002). "evd: Extreme Value Distributions". In: *R News* 2.2. URL: https://CRAN.R-project.org/doc/Rnews/.

Huang, Whitney K. et al. (2019). "New exploratory tools for extremal dependence: χ networks and annual extremal networks". English. In: *J. Agric. Biol. Environ. Stat.* 24.3, pp. 484–501. ISSN: 1085-7117. DOI: 10.1007/s13253-019-00356-4.

Davison, A. C. (2021). *Risk, Rare Events and Extremes*. EPFL.

Capéraà, P., A.-L. Fougères, and C. Genest (1997). "A nonparametric estimation procedure for bivariate extreme value copulas." English. In: *Biometrika* 84.3, pp. 567–577. ISSN: 0006-3444. DOI: 10.1093/biomet/84.3.567.

Alouini, Sonia (2022). *R function to compute a non-parametric estimator of the Pickands function*. Unreleased - Personal communication.

Semadeni, Claudio Andri (2020). "Inference on the Angular Distribution of Extremes". PhD thesis. EPFL.

Cooley, Daniel and Emeric Thibaud (2019). "Decompositions of dependence for high-dimensional extremes". English. In: *Biometrika* 106.3, pp. 587–604. ISSN: 0006-3444. DOI: 10.1093/biomet/asz028.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Janßen, Anja and Phyllis Wan (2020). "k-means clustering of extremes". English. In: *Electron. J. Stat.* 14.1, pp. 1211–1233. ISSN: 1935-7524. DOI: 10.1214/20-EJS1689.

Heffernan, Janet E. and Jonathan A. Tawn (2004). "A conditional approach for multivariate extreme values. (With discussion)". English. In: *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 66.3, pp. 497–546. ISSN: 1369-7412. DOI: 10.1111/j.1467-9868.2004.02050.x.

Southworth, Harry, Janet E. Heffernan, and Paul D. Metcalfe (2020). *texmex: Statistical modelling of extreme values*. R package version 2.4.8.

Keef, Caroline, Ioannis Papastathopoulos, and Jonathan A. Tawn (2013). "Estimation of the conditional distribution of a multivariate variable given that one of its components is large: additional constraints for the Heffernan and Tawn model". English. In: *J. Multivariate Anal.* 115, pp. 396–404. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2012.10.012.