Portfolio formation for a well-diversified portfolio based on graphical models and spherical k-means clustering

Econometrics and Management Science Quantitative Finance

Thesis

Author:
Anne van Oosterhout
(454965)

Supervisor:
Dr. SOR Lonn
Second assessor:
Dr. M Grith

July 27, 2022

Abstract

This paper examines how assets can be selected based on independence structures for extreme values to form a well-diversified portfolio. The independence structures are found through two approaches: Hüsler-Reiss graphical model based on Lasso regularisation and spherical k-means. Both methods confirm that the assets in the same industry (group) are often dependent when extreme values occur. In addition, we find that the diversification level and performance of the portfolio constructed with peripheral assets for non-extreme times, are better compared to the other data-driven portfolios. Nevertheless, we do not find that the difference in performance is significant. In addition, it is evaluated whether the performance of the portfolios resulting from the graphical model is robust regarding the estimation of the regularisation parameter. It is found that their composition changes slightly and their performance does not change significantly.

Contents

1 Introduction

The objective of investors is to make a high return on their portfolio while they consider the risk that they are willing to take. Investors can incorporate risk by diversifying their portfolio by increasing the assets that they hold. Holding multiple assets, a shock in a risk factor may not impact the returns of all assets in a similar way and with the same magnitude. In addition, a positive mean excess return is observed for many equity portfolios as a result of holding more than one asset, i.e. because of diversification (Bessembinder, 2018).

Despite diversifying a portfolio, asset returns are generally influenced by the same risk factors (Fama and French, 1993). In addition, investors may want to take into account how returns behave during a crisis. In this paper, we seek to compose a well-diversified portfolio that is still resistant in times of extreme negative returns to avoid abrupt decrease in value. Diversification is determined by the number of assets on the one hand and their weights on the other hand. One may want to build a portfolio that also performs acceptably on a rainy day. To find which assets an investor should incorporate in this portfolio, she/he can examine the extremal dependency structure between assets. Here, two methods are used to evaluate the dependence structures of the tail of losses.

The first way to measure this dependence is by means of a graphical network. A network consists of nodes, the companies, which are connected through edges, their dependencies. To illustrate, when company A has few edges with low dependencies with other companies, this may indicate that it faces different risk sources than other companies. As a result, company A should be a good addition to diversify the investor's portfolio. On the contrary, when company B has many edges with high dependencies with its neighbouring companies, the impact on the diversification of the portfolio will be relatively small when the investor already holds companies that company B is connected to.

Engelke and Hitz (2020) propose a network for extreme observations that is based on a theory of conditional independence for multivariate Pareto distributions. This method results in a graphical model that allows sparsity for extreme return data. Whereas Engelke and Hitz (2020) find a sparse variogram through block-wise estimation of a minimum spanning tree, this paper imposes sparsity in the underlying dependency structure by means of the data-driven graphical lasso regularisation. A data-driven method is especially useful when estimating for large data sets as it is more time efficient. To evaluate the resulting graphical network and dependencies of the assets, the metrics closeness centrality and betweenness centrality are used to select assets for a central and peripheral portfolio.

The portfolios resulting from the parametric graphical model are compared to portfolios that follow from non-parametric spherical k-means clustering, which is the second method that is evaluated in this paper to determine tail dependence structures. This unsupervised learning method is a good candidate for pattern detection in high dimensional data. Janßen and Wan (2020) propose a clustering method to reduce the complexity of extremal observations. When the assets are split into clusters, a portfolio is created by assigning weights to the assets based on the cluster they are in. Compared to a graphical model, the independence structure of spherical k-means is less informative. The latter only indicates which assets are asymptotically independent of each other whereas a graphical model also gives the level of the dependence when assets are conditionally dependent. In addition, two benchmark portfolios are constructed that do not make use of information regarding the dependence structure of the data. The evaluation of the resulting portfolios is based on their out-of-sample diversification level and performance measured by Sharpe ratio.

The contribution of this paper is twofold. The existing literature does not estimate the extremal conditional dependency structure in a graph in a data-driven way. Engelke and Hitz (2020) suggest that this is a direction of high practical potential and they are already researching this extension themselves, however their research has not yet been published. Moreover, this paper extends the literature by proposing methods to make portfolio management decisions based on extreme returns using graphical models and k-means clustering. There is existing research that uses graphical models for portfolio decisions, however these networks are mostly based on correlation and do not investigate extreme returns (Peralta and Zareei, 2016; Onnela et al., 2002).

We find that the graphical model approach confirms that the assets in the same industry (group) are often dependent when extreme values occur. In addition, we find that the portfolio containing assets that are in the periphery of the Hüsler-Reiss graphical model is well-diversified compared to the portfolios that were constructed based on extremal independence. Concerning performance, there is no portfolio that statistically outperforms the others.

The remainder of this paper is organised as follows. Section 2 discusses the existing literature. Thereafter, the data that is used for the empirical analysis is introduced in Section 3. The methodological approach and evaluation metrics of the graphical models and k-means clustering, together with the construction of the portfolios, are outlined in Section 4. In Section 5, an overview of the results is presented. Finally, Section 6 concludes this paper.

2 Literature

How investors can optimally allocate their wealth across a set of assets originated in Markowitz' mean-variance portfolio and is still the foundation for modern portfolio theory (Markowitz, 1952). Nevertheless, a portfolio that is constructed using sample moments often involves extreme positions, which may not be favourable when constructing a diversified portfolio (Green and Hollifield, 1992). Imposing additional constraints on the weights such that they cannot be negative can reduce the risk in the resulting Markowitz portfolio even when those constraints are wrong (Jagannathan and Ma, 2003). But, mean-variance efficient portfolios are often outperformed by the less complex equally weighted portfolio (DeMiguel et al., 2009). The equally weighted portfolio presents a simple way of diversification and will be used in the formation of the portfolios in this paper.

Portfolio management decisions can be based on network structures. The current literature mostly focuses on correlation based networks (Onnela et al., 2002; Pozzi et al., 2013; Clemente et al., 2021; Zaheer et al., 2022). Onnela et al. (2002) find that the optimal Markowitz portfolio has larger weights in assets that are in the peripherals of a graphical tree. This means that peripheral assets empower low volatility in a portfolio. Moreover, Peralta and Zareei (2016) sustain this empirical finding by providing a link between the Markowitz framework and the financial network, proving that there is a negative relationship between the optimal weights under the Markowitz framework and the centrality of an asset in a correlation based network. Furthermore, they show that a network is a useful device to improve the portfolio selection process by targeting a group of assets according to their centrality.

Extreme value theory (EVT) is concerned with the asymptotic distribution of extreme events. These extreme events happen with low frequency and have a very large impact compared to other observations (Rocco, 2014). The models that have been derived to analyse these events, extrapolate beyond the existing data range and estimate returns for periods of events that have never yet been observed (Engelke and Ivanovs, 2021). For portfolio selection, it is important to consider tail events as they have a large effect on the portfolio's performance. Mainik et al. (2015) propose a portfolio optimisation strategy based on an Extreme Risk Index, which uses EVT to minimise the probability of large losses. They find that their strategy outperforms the minimum-variance and the equally weighted portfolio for assets with heavy tails.

Another challenge in portfolio management is when the number of assets increases. As a result, the total number of parameters to be estimated increases quadratically and the resulting sample covariance matrix is prone to error, which will lead to an incorrect optimisation of port-

folio weights (Michaud, 1989). To decrease the number of parameters, alternative estimators or shrinkage can be used to get a better estimate (Ledoit and Wolf, 2004, 2012). Another way of reducing the number of estimated parameters is through the notion of sparsity. Sparse structures for multivariate extremes can be achieved by means of graphical modelling and clustering (Engelke and Ivanovs, 2021).

The graphical models that are used in this paper to generate sparse structures build further on the approach of Engelke and Hitz (2020), who introduce a theory of conditional independence that uses the perspective of threshold exceedances. This notion applies to multivariate Pareto distributions by combining graphical models and extreme value theory. The advantage of this method is that it produces a simple and sparse structure. In addition, graphs are conveniently interpretative.

The Pareto distribution that Engelke and Hitz (2020) use is the Hüsler-Reiss distribution, characterised by a variogram matrix which can be seen analogously to a covariance matrix of a Gaussian distribution, but then in the world of extremes (Engelke and Hitz, 2020). They impose sparsity in this matrix by finding the underlying minimum spanning tree. This method requires a relatively simple tree to estimate the parameters of the variogram in a block-wise manner, whereas the real structure may be more complex. In addition, it is based on a maximum likelihood estimation which can become computationally expensive when the graph increases.

Alternatively, graphical lasso can be used to disclose the underlying structure and allows for a more advanced network by building a graph via a data-driven approach. Friedman et al. (2008) estimate a sparse inverse covariance matrix using an ℓ_1-penalty through a simple and fast algorithm. However, during the iterations of the algorithm of Friedman et al. (2008), the covariance matrix is estimated, but this method does not guarantee that the covariance matrix is positive definite for every step along the iterations. Nevertheless, in convergence it is positive definite, thus convergence is essential here. In addition, solving the dual problem does not necessarily secure a sparse precision matrix. Mazumder and Hastie (2012) identify this problem and develop a primal approach where the precision matrix is the optimisation target that secures sparsity and positive definiteness of the precision matrix even when convergence has not yet been reached. In the current research, the inverse of the variogram matrix is estimated with graphical lasso and Section 4.3 describes how we make sure that this matrix is invertible.

Sparse structures can also be found by means of spherical k-means clustering. This approach is able to identify groups of concomitant extremes in high dimensions effectively (Fomichov and Ivanovs, 2022). Spherical k-means clustering, similar to other clustering techniques, groups

objects together such that the intra-cluster similarity is maximised while minimising the inter-cluster similarity (Babu et al., 2012). The resulting centres of each estimated cluster characterise the underlying structure of the data. A cluster with one asset entails asymptotic independence, while clusters with multiple assets indicate asymptotic dependence (Janßen and Wan, 2020). León et al. (2017) use this principle to build diversified portfolios using correlation-based clustering. They construct a portfolio by selecting the best asset within each cluster based on Sharpe ratio. Their cluster based portfolios show lower volatility compared to the mean variance portfolio and they conclude that this attributes to the stability and diversity of a portfolio.

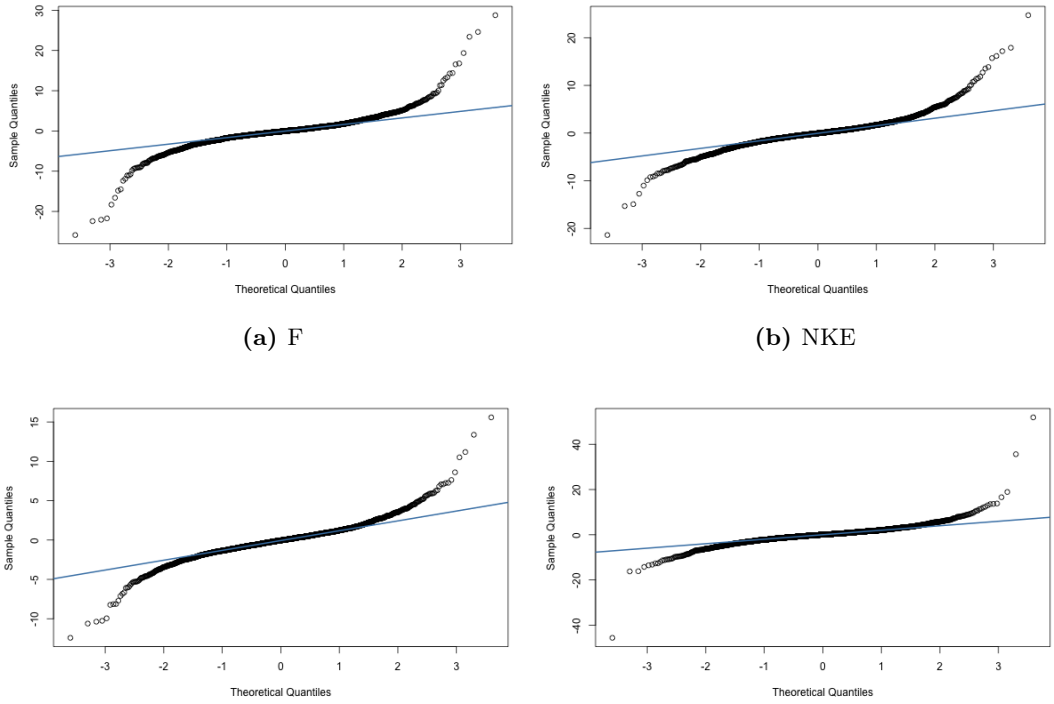
3 Data

The data that is be used in the empirical analysis is daily closing price data of $N = 72$ American companies ranging from December 30$^{\text{st}}$ 2005 until April 1$^{\text{st}}$ 2022 which results in 4092 observations. For estimation and portfolio selection, the in-sample data until April 1$^{\text{st}}$ 2018 is used ($T_{in} = 3082$ observations). To evaluate the portfolios, we work with data of the final four years ($T_{out} = 1010$ observations). The companies are selected based on industry group and all companies have return data in the before-mentioned time horizon. The Global Industry Classification Standard (GICS) identifies 24 industry groups. Of each group, three companies are selected, two large cap companies and one small cap company. The first large cap company in each industry group has a market capitalisation of around 100B$. The second large cap company has a market cap of around 20B$. The small cap companies have a market cap of around 1B$. Information on the companies, categorised by industry group, can be found in Appendix A. Data is retrieved from the database Refinitiv Eikon.

The closing prices of the assets are transformed into log returns by taking the log of the differences and multiplying them by 100 as follows

$$R_t = 100 * \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{1}$$

for $t = 1, \ldots, T$, where T is the total number of return observations. Financial return data often exhibit excess kurtosis, therefore a normal distribution cannot be assumed. This can also be seen in the Q-Q plots in Figure 1, which show the in-sample returns of four randomly chosen assets. When the data would follow a Gaussian distribution, the returns should follow the straight line. However, the returns in Figure 1 have an s-shape, which implies that they are not normally distributed. To verify this, a Jarque-Bera test is performed. Table 1 shows that the test statistics are very high indicating that normality cannot be assumed at any reasonable significance level.



(a) F **(b)** NKE

(c) STE **(d)** IMAX

Figure 1: Q-Q plots of returns of four assets, namely, XOM, F, NKE and IMAX.

Table 1: Jarque-Bera test results to confirm non-normality for four asset returns

	JB-statistic
F	45124
NKE	16396
STE	9583
IMAX	279057

4 Methodology

In this section, the methodological approach to construct well-diversified portfolios is explained. First, some theoretical background regarding multivariate extreme value theory and conditional independence is described. Thereafter, the method to estimate a sparse graph is discussed and it is outlined how the metrics for graph evaluation will be the basis for portfolio selection. Subsequently, the method for spherical k-means clustering is explained and how a diversified portfolio may result from this method. Finally, the construction of the portfolios is described in detail and the evaluation measures regarding diversification and performance are presented.

4.1 Data preparation

A fundamental assumption in EVT is that the data is independent and identically distributed. In order to obtain this, the approach of Engelke and Volgushev (2020) is followed, who remove temporal dependence. This is necessary as with the expansion of companies, often their stock price grows too. As a result, the price of the stock may be very different at the beginning of the period compared to the end. In addition, the returns as calculated in Equation 1 are multiplied by -1 as we are interested in the left tail for the analysis. The serial dependence is removed by filtering the univariate series through ARMA-GARCH processes, which is proposed by Hilal et al. (2014). Models with various numbers of parameters are compared by their AIC. An ARMA(p,q)-GARCH(r,s) model for a univariate series of company log losses (L_t) for every company can be formulated as follows

$$\begin{aligned} L_t &= \mu_t + \sigma_t X_t \\ \mu_t &= \mu + \sum_{i=1}^{p} \psi_i L_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \\ \sigma_t^2 &= \omega + \sum_{i=1}^{r} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{s} \beta_i \sigma_{t-i}^2 \\ \varepsilon_t &= L_t - \mu_t = \sigma_t X_t \\ X_t &\sim i.i.d. \end{aligned} \tag{2}$$

where $|\psi_i| < 1$ and $|\theta_i| < 1$ for stationarity of the conditional mean. In addition, for the GARCH part of the model, we assume that $\omega > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ and $\sum_{i=1}^{r} \alpha_i + \sum_{i=1}^{s} \beta_i < 1$ to guarantee that σ_t^2 is positive and stationary. The parameters in the model are estimated by maximum likelihood such that no distribution on the shocks X_t needs to be assumed. The resulting filtered losses of a company j are then

$$X_t^{(j)} = \frac{L_t^{(j)} - \hat{\mu}_t^{(j)}}{\hat{\sigma}_t^{(j)}} \tag{3}$$

where $\hat{\mu}^{(j)}$ and $\hat{\sigma}^{(j)}$ are the estimated mean and standard deviation of the model estimated in Equation 2. The data $\mathbf{X} = (X_1, \ldots, X_N)$ is approximately independent and identically distributed and their tail dependence can be modelled using an extreme graphical model or spherical k-means clustering.

4.2 Multivariate extreme value theory

Extreme value theory is about modelling the tail of a distribution. In a multivariate case, the model also takes into account the dependence in the univariate extremes. The tail behaviour of

the random vector $\mathbf{X}$ can be described through two methods: the component-wise maxima and the threshold exceedances approach. The modelling strategy that is used to study multivariate dependencies in graphical models is the threshold approach (Engelke and Hitz, 2020). For this method, it is common to standardise the data to multivariate Pareto scale. The exceedances of data $\mathbf{X}$ over a high threshold $u \to \infty$ converges to limit $\mathbf{Y}$ (Resnick, 2007),

$$\mathbb{P}(\mathbf{Y} \leq \mathbf{z}) = \lim_{u \to \infty} \mathbb{P}\left(\frac{\mathbf{X}}{u} \leq \mathbf{z} \middle| \|\mathbf{X}\|_\infty > u \right), \tag{4}$$

such that the random vector $\mathbf{Y}$ is multivariate Pareto distributed on the L-shaped space $\mathcal{L} = \{x \geq 0, \|x\|_\infty \geq 1\}$. Here, the assumption is made that $\mathbf{Y}$ has a positive and continuous density $f_{\mathbf{Y}}$ on $\mathcal{L}$. The density of $\mathbf{Y}$ is then

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{\lambda(\mathbf{y})}{\Lambda(\mathbf{1})}, \tag{5}$$

where Λ is the exponent measure of the max-stable distribution (Engelke and Hitz, 2020), $\mathbf{1}$ is a vector of ones and $\Lambda(\mathbf{1})$ is a normalisation constant. Thus, the density $f_{\mathbf{Y}}$ is proportional to the density $\lambda(\mathbf{y})$ of the exponent measure Λ. In practice, the columns of the data are first transformed to standard Pareto distributions by means of

$$\tilde{\mathbf{X}}_{\mathbf{j}} = \frac{1}{1 - F(\mathbf{X}_j)}, \tag{6}$$

where $F(\mathbf{X}_j)$ is the uniform distribution of asset j. Then the data of the time points that contain no extreme values are disregarded from the data, i.e. there is no extreme return in observation i if for every asset j in $\tilde{\mathbf{X}}_{ij}$ is smaller than the quantile threshold. Finally, the remaining observations are standardised by dividing them by the quantile threshold. The extremes in the resulting data $\tilde{\mathbf{X}}$ then have a value higher than 1 and non extreme returns are smaller than 1.

Because some components of $\mathbf{X}$ may not have converged to the limiting distribution of $\mathbf{Y}$, it has become common practice to censor the data when estimating the likelihood in order to avoid biased estimates of the parameters (Engelke and Hitz, 2020). This censoring entails that when there is a data point $\tilde{\mathbf{X}}$ that exceeds a high threshold, simultaneously there may be data points that do not exceed this threshold. The latter points are censored such that their value is set to 1. In this way, the information that is used about these points entails the fact that they are smaller than the threshold instead of their exact value.

4.3 Graphical model for extremes

4.3.1 Conditional independence for graphical models

A graph $\mathcal{G} = (V, E)$ is composed of a nodes set $V = \{1,\ldots,d\}$, also called vertices, and a set of edges $E \subset V \times V \setminus (i, i)$ which connects the vertices. In the case of our data, the vertices represent the assets and the edges their dependence. A (sub)set of nodes of which all vertices are connected such that $(i, j) \in E$ is called complete.

The notion of conditional independence for multivariate Pareto distributions is necessary to reduce the number of parameters to be estimated. Engelke and Hitz (2020) exploit this notion in extreme value theory. As it is only defined on product spaces, they come up with a new insight of conditional independence for a multivariate Pareto distribution and restrict $\mathbf{Y}$ to product spaces. They show this by considering $\mathbf{Y}^j$, which is $\mathbf{Y}$ conditioned on the event that $\{Y_j > 1\}$ for any vertex $j \in V$. The resulting $\mathbf{Y}^j$ has support on a product space.

Given that random variable $\mathbf{Y}$ follows a multivariate Pareto distribution and has positive and continuous density, and $A, B, C \subset V$ are non-empty disjoint subsets whose union is V. When $\mathbf{Y}_A$ is conditionally independent of $\mathbf{Y}_C$ given $\mathbf{Y}_B$, it can be written as

$$\mathbf{Y}_A \perp_e \mathbf{Y}_C | \mathbf{Y}_B. \tag{7}$$

For graphical models, conditional dependence means that vertices are connected. Contrary, when two vertices i and j in V are not connected in graph $\mathcal{G}$, then Y_i and Y_j are said to be conditionally independent. For conditional independence there are two underlying assumptions. One is that $\mathbf{Y}$ has a density and the second assumption is that there is no asymptotic independence. When one asset would be asymptotically independent of all other assets, the graph would become disconnected.

4.3.2 Graphical models and Hüsler-Reiss distribution

The graphical model $\mathcal{G} = (V, E)$ can be identified by a Hüsler-Reiss Pareto distribution. The density of the exponent measure for every root node $r = 1, \ldots, N$, i.e. every asset, is

$$\lambda(\mathbf{y}) = y_r^{-2} \prod_{i \neq r} y_i^{-1} \phi_{N-1}\left(\tilde{\mathbf{y}}_{\setminus r}; \Sigma^{(r)}\right), \tag{8}$$

where $\tilde{\mathbf{y}}_i = \frac{log(y_i)}{log(y_r)} + \Gamma_{ir}/2$. In addition, $\phi_{N-1}(\cdot; \Sigma^{(r)})$ is a $(N-1)$-dimensional centred normal distribution with covariance matrix $\Sigma^{(r)}$, which is constructed as follows

$$\Sigma^{(r)} = \frac{1}{2}\left\{\Gamma_{ir} + \Gamma_{jr} - \Gamma_{ij}\right\}_{i,j \neq r} \in \mathbb{R}^{(N-1)\times(N-1)}. \tag{9}$$

The $\mathbb{R}^{(N-1)\times(N-1)}$ covariance matrix $\Sigma^{(r)}$ is strictly positive definite and contains the graphical structure of a Hüsler–Reiss Pareto distribution. The dimension is $(N-1)$ because the row and column of the root node r are omitted. Although r is fixed, the distribution does not depend on the arbitrary choice of r. The distribution has the advantageous characteristic of stable marginals and is therefore a good candidate for graphical models. These marginals are required for the density of the exponent measure.

To estimate the covariance matrix, we need the negative definite variogram matrix Γ $(N \times N)$, which is symmetric and has zeros on its diagonal. The empirical $\hat{\Gamma}$ matrix can be estimated by averaging the $\hat{\Gamma}^{(r)}$ matrices for every root node r. For a root node, the $\Gamma^{(r)}$ matrix is created out of a special covariance matrix $(\Sigma_{ex}^{(r)})$. This latter matrix is formed using the log of the observations when asset r exceeds a threshold. The relationship between $\Gamma^{(r)}$ and $\Sigma_{ex}^{(r)}$ is as follows

$$\begin{aligned} \hat{\Gamma} &= \frac{1}{N}\sum_{r=1}^{N}\Gamma^{(r)}, \\ \hat{\Gamma}^{(r)} &= \mathbf{1}\ \mathrm{diag}(\Sigma_{ex}^{(r)})^T + \ \mathrm{diag}(\Sigma_{ex}^{(r)})\mathbf{1}^T - 2\Sigma_{ex}^{(r)} \\ \Sigma_{ex}^{(r)} &= \mathrm{Var}(\ \log Y | Y_{ir} > 1, \text{ for } i = 1, \ldots, T) \end{aligned} \tag{10}$$

The variogram matrix can only be estimated if there is no asymptotic independence, i.e. the graph must be connected. Using the estimate, the covariance matrix can be estimated as in Equation 9. The inverse of this covariance matrix, $\Theta^{(r)} = (\Sigma^{(r)})^{-1}$, also known as the precision matrix, determines the graph structure. It then holds for $i, j \in V$ with $i \neq j$ and for any $r \in V$

$$Y_i \perp_e Y_j | Y_{V\setminus\{i,j\}} \iff \begin{cases} \Theta_{ij}^{(r)} = 0 & \text{if} \quad i,j \neq r \\ \sum_{l\neq} \Theta_{lj}^{(r)} = 0 & \text{if} \quad i = r, j \neq r, \\ \sum_{l\neq r} \Theta_{il}^{(r)} = 0 & \text{if} \quad j = r, i \neq r. \end{cases} \tag{11}$$

Thus, the vertices i and j are not connected, i.e. conditionally independent, if there is a zero in the $(ij)^{th}$ element of the precision matrix, given that the precision matrix is estimated with a root node that is not equal to either node i or j. Alternatively, there is no connection between

i and j if the sum of the row or column is equal to zero when one of them is equal to the root node.

4.3.3 Graphs based on ℓ_1-regularisation

When the graph is known and decomposable, the Hüsler-Reiss distribution can be estimated by means of the likelihood of the cliques (Engelke and Hitz, 2020). However, such a graph structure is mostly not known beforehand, especially in risk analysis of financial networks. In these networks, there is no explicit spatial structure between institutions. One way to estimate the graph structure is by means of minimum spanning trees, as is done by Engelke and Hitz (2020). A downside of the approach is that trees are somewhat simplistic structures and some information may get lost due to this. In addition, this method becomes computationally expensive when the data set is large ($N > 40$). Alternatively, a more general underlying structure can be learned in a data-driven way.

To estimate a sparse graph, Engelke and Hitz (2020) suggest in their discussion to use graphical lasso for extremes as an alternative to their block-wise estimation method. The approach would be similar to graphical lasso of Friedman et al. (2008), however the main difference between the two methods is that extremal graphical lasso makes use of empirical estimates of Γ and $\Sigma^{(r)}$, whereas the original method makes use of the empirical covariance matrix of a multivariate normal distribution.

With the empirical $\hat{\Sigma}^{(r)}$ and $\hat{\Gamma}$ of Equations 9 and 10, for some root node $r \in V$, extremal graphical Lasso can be performed. For a tuning parameter ρ, this is equivalent to solving the convex problem

$$\hat{\Theta}_\rho^{(r)} = \underset{\Theta \succeq 0}{\arg\max} \quad \log(\det(\Theta)) - \operatorname{tr}(\hat{\Sigma}^{(r)}\Theta) - \rho \sum\nolimits_{i \neq j, i, j \neq r} |\Theta_{ij}| \,. \tag{12}$$

The optimisation is repeated N times, such that every vertex is once the root node. This will result in N graphs structures for every value for ρ. The overall graph structure for each ρ can be extracted from the extremal graphical lasso estimates by means of a majority vote, i.e. if there is an edge between vertices i and j in the majority of the N graph estimates, this edge is also present in the final graph structure. Mathematically this results in

$$(i,j) \in \hat{E}_\rho \iff \frac{1}{N-2} \#\{r \in V \backslash \{i,j\} : (\hat{\Theta}_\rho^{(r)})_{ij} \neq 0\} \geq \frac{1}{2}. \tag{13}$$

When the graph structure is discovered, the variogram matrix and covariance matrix can be estimated as a second step. The optimal value for ρ is chosen by evaluating the graph for each

ρ by means of the Bayesian Information Criterion (BIC). This model selection criterion has the following formula

$$\text{BIC} = |E| \cdot T - 2 \cdot \log(L) \tag{14}$$

where $|E|$ is the cardinality of the edges set, i.e. the number of edges in a graph, T is the number of in-sample observations and L is the censored likelihood estimate of the graph

$$f(\mathbf{y}, \Gamma) = \frac{1}{\Lambda(\mathbf{1}, \Gamma)} \prod_{i=1}^{T} \lambda(\mathbf{y}, \Gamma), \tag{15}$$

where $\Lambda(\mathbf{1}, \Gamma)$ is the exponent measure and $\lambda(\mathbf{y}, \Gamma)$ the density of the exponent measure.

4.3.4 Graph evaluation

When a graph has been estimated, it is analysed in more depth such that portfolio decisions can be made based on the graph. First of all, more or less sparse graphs are estimated as a result of a varying parameter ρ. Where a higher value of ρ results in a more sparse graph. Having a graph for every ρ, they can be compared based on their log likelihood to find the optimal graphical structure and its corresponding tuning parameter.

Given the best graph, individual nodes are examined more closely. Popular measures for graphs are closeness and degree centrality (Freeman, 1977). Closeness centrality is the sum of the shortest distance in the graph, $d(i, j)$, between node i and j measured through correlation (Mantegna, 1999). This distance can be retrieved by weighting the edges of the graph by their correlation. The measure reflects the importance of node i relative to the others and can be computed as

$$C_c(i) = \frac{N - 1}{\sum_{i=1}^{N} d(i, j)}. \tag{16}$$

A central asset is likely to have a short distance to all other assets. Since closeness centrality is defined by the inverse of the distance between the assets multiplied by a constant, a central asset has a relatively high closeness centrality level. In addition, degree centrality measures the number of edges connected to a node, or the number of direct neighbours a node has. Mathematically degree centrality is

$$C_d(i) = \frac{\deg(i)}{N - 1}, \tag{17}$$

where $N - 1$ normalises the degree. An asset that is connected to many other assets, i.e. its degree centrality measure is relatively high, is characterised as a central asset. Assets that have a high degree are more likely to have a shorter distance to all the other assets. Yet, the two measures are both considered because not all information regarding centrality is captured in

either. To find which assets are the most central and which are the least central, we create two rankings of the assets based on the separate measures. Their final centrality rank is the average of the previous two.

4.3.5 Robustness analysis

With the ℓ_1-regularisation, a graph can be found for optimal ρ. For this value of ρ, the graph is theoretically optimal, however this graph may not reflect what exactly happens in reality. The optimal ρ is based on the BIC, which in general favours more parsimonious models over more complex models. Yet, an edge between two assets may be important even if it is small and will disappear with higher values of the regularisation parameter. Therefore, as a robustness analysis, we will construct graphs with a higher and lower value of the regularisation parameter as well. Evaluating these graphs gives us the opportunity to review whether central and non-central assets are also categorised as such in more and less dense graphs. Furthermore, the performance of the central and peripheral portfolios will be determined again and it is shown what impact on performance is when the portfolio selection is based on non-optimal graphs.

4.4 Spherical k-means clustering

An alternative method to estimate the dependence between extreme observations, or rather their independence, is k-means clustering. This machine learning method partitions data in a predefined number k clusters where each asset is assigned to a cluster. The clusters can be seen as subsets holding observations such that the observations in one subset are more similar to each other than to observations in other subsets. For k-means clustering, the optimal solution is found by minimising the distance between the observations $\mathbf{Y} = (\mathbf{y}_1, ..., \mathbf{y}_N)^T$ and a set of cluster centres or prototypes, $P = \{\mathbf{p}_1, ..., \mathbf{p}_k\}$,

$$W(P, S) := \int_{\mathbb{R}^d} \min_{\mathbf{p} \in P} d(\mathbf{y}, \mathbf{p}) S(d\mathbf{y}) \in [0, \infty], \tag{18}$$

where S is a probability measure (Janßen and Wan, 2020). Often the squared Euclidean distance is used for $d(\mathbf{y}, \mathbf{p})$, however when data is multivariate it can be interesting to incorporate the angles between the observations such that more information regarding the data can be used. As there are 72 assets in the data set for the empirical evaluation, we will continue with the spherical k-means. This method finds clusters based on cosine dissimilarity and prototype-based partitioning, and is therefore able to include these angles by means of

$$d(\mathbf{y}, \mathbf{p}) = 1 - \cos(\mathbf{y}, \mathbf{p}) = 1 - \frac{\langle \mathbf{y}, \mathbf{p} \rangle}{\|\mathbf{y}\| \|\mathbf{p}\|}. \tag{19}$$

It is also standard approach to standardise the observations when using empirical data when using this method. Similar to the data used to estimate the graphs, the data is standardised to have standard Pareto marginals. In addition, the data is normalised to one such that the magnitude of the observations is less influential for the optimisation.

In the `skmeans` R-package, there are multiple built-in methods for computing the spherical k-means partitions. In this research, the `pclust` method is used, which iterates between determining optimal memberships for fixed prototypes, and computing optimal prototypes for fixed memberships. In addition, we make use of hard partitions, meaning that an asset can only belong to one cluster at a time.

4.5 Portfolio formation

Now that we have explained two methods to identify the dependence structure of assets for extreme values, a way to transform this information into portfolio selection is required. Below, it is described how benchmark portfolios are formed. In addition, the portfolio formation methods for graphical model and spherical k-means and their corresponding weights are explained.

4.5.1 Benchmark portfolios

To compare the portfolios based on complex models, two benchmark portfolios are formed without making use of the information from the dependence structures of the assets. The first portfolio is constructed by combining all 72 assets in an equally weighted portfolio. The choice for a simple $1/N$ portfolio is that its out-of-sample Sharpe ratio outperforms various sample-based mean-variance strategies (DeMiguel et al., 2009).

In addition, a second portfolio is created to have maximum diversification. Whereas Choueifaty and Coignard (2008) find a maximum diversification portfolio as a result of establishing the weights based on the diversification ratio (which is explained below in Section 4.6.1), here the asset selection is based on this ratio together with the portfolio's standard deviation of the marginal risk contribution. The reason for this approach is that this portfolio is built to hold 24 assets with equal weights. Thus, the maximum diversification portfolio here does not seek weights such that there is maximum diversification, but tries to find the combination of 24 assets in an equally weighted portfolio that maximises diversification based on the two measures.

Testing all combinations of 24 out of the 72 assets entails evaluating 7.95E^{18} portfolios, which is very computationally expensive. Therefore, the maximum diversification portfolio is approximated by finding a combination of 24 assets out of 100,000 possible portfolios, which is most diversified for in-sample data according to the measures described in Section 4.6.1. As a result,

the most diversified portfolio is the combination of assets with the highest diversification ratio and lowest standard deviation of the marginal risk contribution. To find this portfolio, we do a double sort on these two measures as they may not be of the same magnitude. For the out-of-sample evaluation, the 24 assets are again combined in an equally weighted portfolio. The assets that make up this portfolio can be found in Appendix B.

4.5.2 Portfolio based on graphical model

Given the optimal graphical model, two portfolios containing assets based on their centrality are constructed. For these portfolios, we need to specify the number of assets contained by them. One condition of these portfolios is that one asset should not be simultaneously categorised as central and non-central. Since the data consists of 72 assets, we choose that the number of assets in the portfolios is 24. In this way, there are still assets that can be defined as neither central nor non-central. In addition, there are also exactly 24 industry groups.

The first portfolio based on the graphical model, the central portfolio, contains assets that are ranked the most central. This portfolio holds assets which are expected to have more dependence in their extreme losses, and therefore, it may perform badly in times of financial distress. To contrast the central portfolio, a peripheral portfolio is created from the assets that have the lowest centrality scores. Onnela et al. (2002) find that combining companies from the peripherals leads to a diversified portfolio. Both portfolios are evaluated for out-of-sample data with equal weights.

4.5.3 Portfolio based on spherical k-means

For the spherical k-means portfolio, first the number of clusters should be determined. Too few clusters would entail that there can still be considerable differences between the assets in one cluster. On the other hand, too many clusters may result in clusters that still depend on each other when extreme values occur. To establish a suitable number of clusters, an elbow plot is built which shows the minimum distance between the observations and the cluster centres for a range of values of k.

Given a value of k, the assets in each cluster hold extremal dependencies. In order to create a diversified portfolio, each cluster has a weight of $\frac{1}{k}$, and the assets within a cluster hold equal weights. To illustrate this, if assets are part of cluster k with a cardinality of $|C_k| = N_k$, the weights of these assets equal $\frac{1}{k} \cdot \frac{1}{N_k}$.

4.6 Portfolio evaluation technique

4.6.1 Portfolio diversification technique

Having a method that presumably produces diversified portfolios, we need to measure diversity of the portfolios, to verify our assumptions. Two measures will be regarded, the diversification ratio (DR) and the marginal risk contribution (MRC). The DR is the ratio of the weighted average of volatilities divided by the portfolio volatility and was introduced by Choueifaty and Coignard (2008). It measures how much the volatility decreases by building a portfolio instead of holding separate assets. The measure is calculated as

$$DR(w) = \frac{\sum_{i=1}^{N} w_i \sigma_i}{\sigma_p}, \tag{20}$$

where w_i is the weight in each asset i, σ_i is the standard deviation of each asset i in the portfolio and σ_p is the volatility of the portfolio. A large value of DR implies a low portfolio volatility. This means that it is enhances diversification by putting these assets together in a portfolio instead of holding the assets separately. In addition, the MRC, $\partial_w \sigma_p$, measures how much each asset in the portfolio contributes to the total risk of the portfolio (Maillard et al., 2010). The MRC is computed as follows

$$\partial_w \sigma_p = \frac{\partial \sigma_p}{\partial w} = \frac{w_i \sigma_i^2 + \sum_{j \neq i} w_j \sigma_{ij}}{\sigma_p}, \tag{21}$$

where σ_i^2 is the variance of asset i and σ_{ij} the covariance between assets i and j. A well-diversified portfolio is one where all assets contribute to the same extent to the total risk. In the analysis, the standard deviation of the marginal risk contributions is evaluated such that a low volatility indicates a well-diversified portfolio.

4.6.2 Portfolio performance technique

In addition to their diversification level, the portfolios are evaluated regarding their performance. The performance metric that is used to compare the portfolios is the Sharpe ratio (SR), which can be interpreted as the portfolio return corrected for volatility. The SR is computed as follows

$$\text{SR} = \frac{\mu_p^{ex}}{\sigma_p} * \sqrt{252} \tag{22}$$

where μ_p^{ex} is the mean of daily excess return of the portfolio and σ_p is the standard deviation of the daily gross returns. The Sharpe ratio is annualised by multiplying it with $\sqrt{252}$. A higher SR can be due to high returns or low volatility. To break this down and find the source of a high

SR, the portfolio mean and variance are also investigated separately. To evaluate whether the Sharpe ratios of one portfolios is significantly higher than that of another portfolio, the Sharpe ratios are tested. The null hypothesis is that the Sharpe ratio of one portfolio is equal to the Sharpe ratio of another (Ledoit and Wolf, 2008). The test statistic is the difference between the Sharpe ratios of two portfolios p_1 and p_2

$$\hat{\Delta} = \widehat{SR}_{p_1} - \widehat{SR}_{p_2} = \frac{\hat{\mu}^{ex}_{p_1}}{\hat{\sigma}_{p_1}} - \frac{\hat{\mu}^{ex}_{p_2}}{\hat{\sigma}_{p_2}}. \tag{23}$$

Through the application of the Delta method, we can find convergence in distribution, namely

$$\sqrt{T}(\hat{\Delta} - \Delta) \xrightarrow{d} \mathrm{N}\left(0, \nabla' f(v) \Psi \nabla f(v)\right) \tag{24}$$

where $\nabla f(v)$ is a gradient. For a consistent estimator $\hat{\Psi}$, of Ψ, we use a heteroskedasticity and auto-correlation robust (HAC) kernel estimation. The kernel that is used here is the commonly used Parzen kernel. HAC inferences are often liberal when applied to small to moderate sample sizes (Ledoit and Wolf, 2008). However, here the Sharpe ratio is computed and tested for the out-of-sample data which consists of 1010 observations, which we believe is sufficiently large. The resulting standard errors, $s(\hat{\Delta})$, for $\hat{\Delta}$ and the p-values are calculated as

$$s(\hat{\Delta}) = \sqrt{\frac{\nabla' f(\hat{v}) \hat{\Psi} \nabla f(\hat{v})}{T}} \qquad\qquad \hat{p} = 2\Phi\left(-\frac{|\hat{\Delta}|}{s(\hat{\Delta})}\right)$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

5 Results

In this section, the empirical results are discussed. First, the results of the graphical model approach are elaborated on and it is investigated which assets will be part of the central and peripheral portfolios. To estimate the graphs, the R package `graphicalExtremes` of Engelke and Hitz (2020) is used. Thereafter, the portfolios based on the k-means clustering approach is outlined. For the implementation of spherical k-means clustering, the R package `skmeans` is utilised. The resulting portfolios are compared regarding the portfolios' diversification and performance for out-of-sample data.

5.1 Generate i.i.d. data

To apply multivariate EVT, the assumption of independent and identically distributed data is required. In order to be able to assume that the data is approximately i.i.d., the ARMA(1,1)-GARCH(1,2) model was applied to the data as it had the lowest AIC score. An overview of various ARMA-GARCH models and their AIC score are attached in Appendix C.

5.2 Graphical model based portfolio

After the i.i.d. data has been transformed to have Pareto margins following Equation 6, the extremal conditional dependence between the assets can be estimated. For a threshold u equal to 90% the number of observations for which at least one asset return exceeds this threshold remains $T = 2715$. Thereafter the empirical variogram is estimated. For every root node k a variogram matrix is estimated using data of the time points where the return of asset k exceeds one. The empirical variogram matrix is calculated as the element-wise average of the k variogram matrices.

The empirical variogram is then used in the graphical lasso method to estimate sparse graphs. In Figure 2 the BIC of graphs for $\rho = (0, 0.005, \ldots, 0.050, 0.055)$ is presented. The figure shows that an optimal graph has a ρ of around 0.050. This is the global minimum since for ρ much larger than 0.055 the graph becomes disconnected and it is not possible to estimate a variogram matrix anymore. In addition, the figure shows that a sparser graph does not necessarily entail a better fit. This phenomenon highlights that some connections between nodes are essential to a better fit. To get a more precise insight into which ρ is optimal, we zoom in and review a smaller

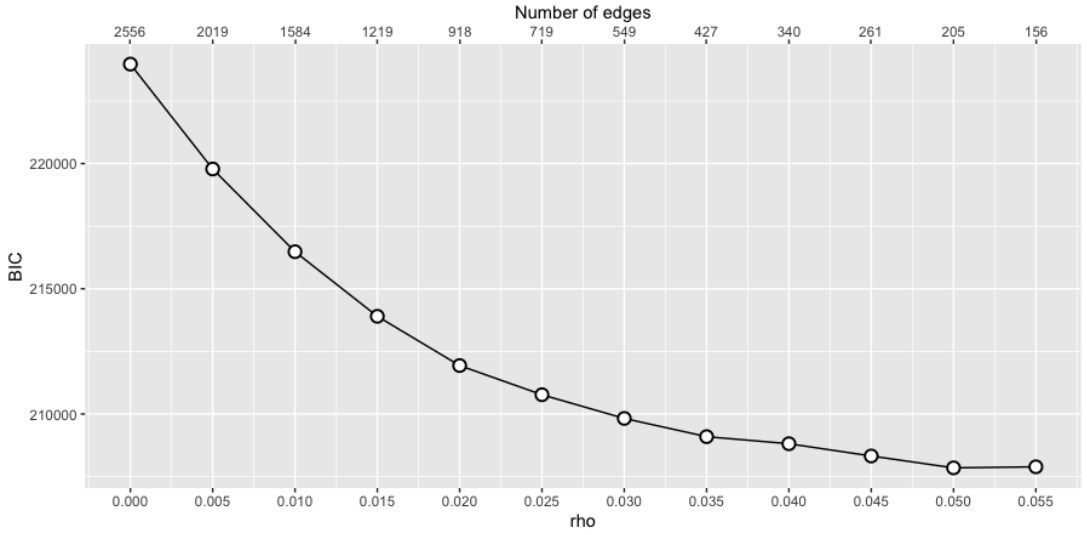


Figure 2: BIC results for a range of different values for tuning parameter ρ and the corresponding number of edges of the graph.

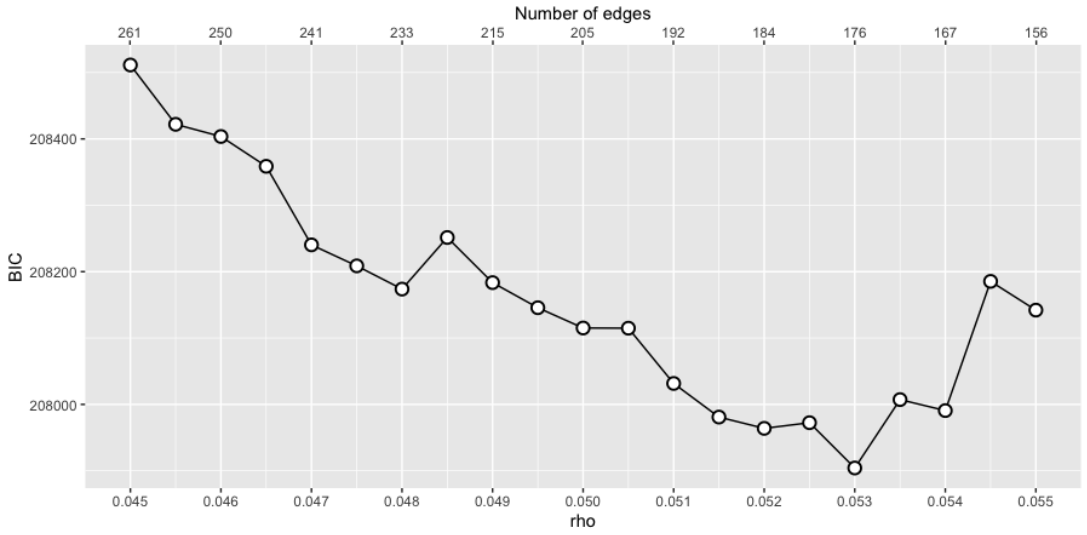


Figure 3: BIC results for specific values for tuning parameter ρ and the corresponding number of edges of the graph.

range of ρ. These results can be seen in Figure 3. The latter figure shows that the optimal value for ρ equals 0.053 with a BIC measure of 207904. The corresponding optimal graph has 176 edges and a figure of this graph can be found in Figure 4. The red and yellow vertices are the central and peripheral assets respectively, the grey vertices do not belong to either of these categories. As one edge connects two assets, on average one asset is connected with 4.88 other assets. In this graph, the most central assets are connected with 10 other assets and the least connected asset has 2 edges.

From the graph that is estimated with the optimal ρ, the closeness and degree centrality are calculated. Recall that a central asset is likely to have a short distance to all other assets and is expected to be connected to many other assets. Since the portfolios that are made contain 24 assets, Table 2, shows the central assets, ranked as 1-24, and the peripheral assets, ranked as 49-72, that will make up the central and peripheral portfolio respectively. The rank of the assets is the result of the average of the rank of the closeness and degree centrality. In Appendix D, the ranks of the two centrality measures are given separately. They are not exactly equivalent to each other in terms of ranking the assets, but they do hold strong similarity. Of the 24 most central assets, the two measures have 17 overlapping assets. This number is 16 out of 24 for the peripheral assets. In addition, non of the assets that is characterised as a central asset is labelled peripheral by the other measure and vice versa.

From Table 2 we learn that stocks that are found to be central are often part of the same industry. Industries that have a lot of central stocks are the consumer staples, financial, materials and utilities industry. In addition, all assets of the industry group telecommunication services are

labelled as central. This result is quite surprising as these sectors provide services or goods that are always in demand. Especially the consumer staples sector companies, a sector that is characterised by relatively low volatile assets, are in the centre of the estimated graph. Thus, we find that in times of crisis the returns of these generally low market beta assets are highly dependent.

The industry pattern is not discernible for the set of peripheral stocks, in which 18 out of the 24 industry groups is represented by at least one asset. Yet, the consumer discretionary

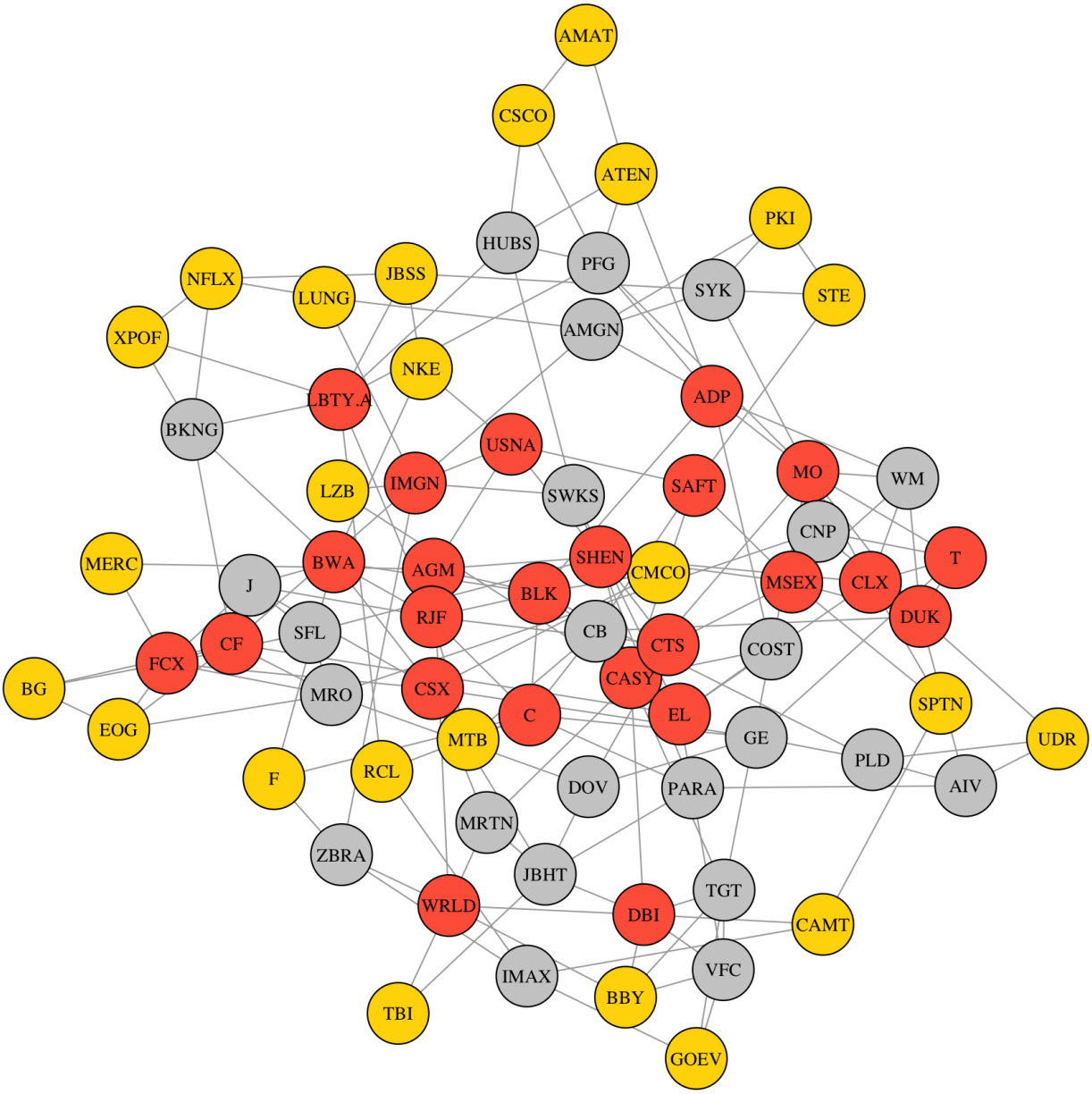


Figure 4: Optimal Hüsler-Reiss graphical model for 72 asset for $\rho = 0.053$, where the red and yellow vertices indicate the central and peripheral assets respectively and the grey vertices are neither part of the central nor of the peripheral portfolio.

industry is represented by seven assets. This is also surprising since companies in this sector are characterised by producing goods or offering services that are non-essential to enjoy basic living conditions. These companies are cyclical, meaning that when the economy is growing, they sell more products and services and if the economy is weakening, these companies sell fewer

Table 2: Lists of assets that are contained in the central and peripheral portfolios

Rank	Central	Size	Industry	Rank	Peripheral	Size	Industry
1 (most central)	MO	L	CST	49	CMCO	S	I
2	SHEN	S	CSE	50	JBSS	S	CST
3	BLK	L	F	51	NFLX	L	CSE
4	RJF	M	F	52	SPTN	S	CST
5	DUK	L	U	53	BBY	M	CD
6	IMGN	S	HC	54	RCL	M	CD
7	CASY	M	CST	55	LZB	S	CD
8	ADP	L	IT	56	ATEN	S	IT
9	CLX	M	CST	57	LUNG	S	HC
10	AGM	S	F	58	CAMT	S	IT
11	C	L	F	59	NKE	L	CD
12	LBTY.A	M	CSE	60	MTB	M	F
13	CF	M	M	61	EOG	L	E
14	MSEX	S	U	62	F	L	CD
15	BWA	M	CD	63	UDR	M	RE
16	DBI	S	CD	64	STE	M	HC
17	CTS	S	IT	65	BG	M	CST
18	CSX	L	I	66	MERC	S	M
19	USNA	S	CST	67	CSCO	L	IT
20	EL	L	CST	68	PKI	M	HC
21	FCX	L	M	69	XPOF	S	CD
22	SAFT	S	F	70	GOEV	S	CD
23	WRLD	S	F	71	TBI	S	I
24	T	L	CSE	72 (least central)	AMAT	L	IT

Note: this table shows the 24 most central assets and the 24 least central assets for a Hüsler-Reiss graphical model estimated with regularisation parameter $\rho = 0.053$. The ranking is the average ranking of the closeness and degree centrality measures. In addition, the market capitalisation is given where S, M and L stand for small, medium and large market cap. Abbreviations of the industry to which the companies belong are given in the industry columns. The 11 industries with their abbreviation between parentheses are as follows: Energy (E), Materials (M), Industrials (I), Consumer discretionary (CD), Consumer staples (CST), Health care (HC), Financials (F), Information technology (IT), Communication services (CSE), Utilities (U) and Real estate (RE).

products. Thus, when there occurs a shock in the market or a crisis, the returns of these assets may represent the decreased demand for these goods and services. However, the returns of these assets do not show this pattern here. In addition, almost half of the assets that make up this set has a small market capitalisation whereas the distribution of small, medium and large market capitalisation is more equally distributed for central stocks. From this we conclude that small firms are often resilient in times when other assets show extreme losses.

5.3 Spherical k-means based portfolio

To create the spherical k-means portfolio, first the number of clusters must be established. Figure 5 shows an elbow plot for different values of k. The minimised average distance between the observations and their cluster centre decreases while the number of centres increases. When the number of clusters would be increased to $k = 72$, it would mean that each cluster holds exactly one asset. In Figure 5, there is no distinct value for k such that adding another cluster would decrease $W(P, S)$ only slightly. Therefore, the analysis is executed for $k = 10$ and $k = 24$ and the resulting clusters are compared.

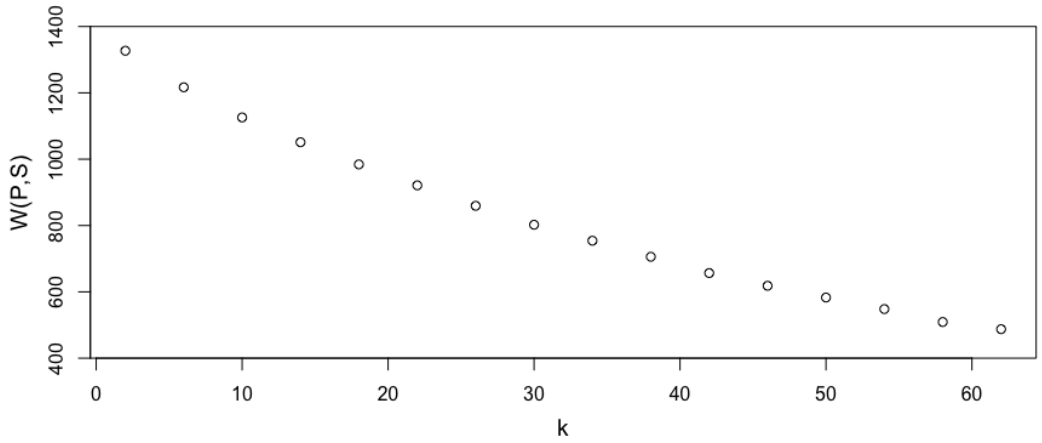


Figure 5: Elbow plot for different values of k.

For $k = 10$ clusters, we can find which assets form the clusters. Based on the combination of assets in a cluster, an underlying characteristic can be uncovered. Figure 6 shows how far the assets are from the centres of the clusters, where the centres are normalised, such that the asset that is closest to a centre has value 1 (white box) and assets that are far from a centre have value 0 (red box). Table 3 gives an overview of which asset belongs to which cluster. The assets in a cluster are asymptotically independent of the assets in other clusters.

When reviewing these results, we see that there are four clusters which only contain one asset.

Three of them hold an asset with a small market capitalisation. Furthermore, some clusters are related to industry (groups). Cluster 2 contains three retailing assets, Cluster 8 contains utilities and real estate assets and Cluster 9 can be characterised by the energy and materials industry groups. Lastly, Cluster 10 is the largest cluster having a cardinality of 49. There may still be some independence between the assets in this last cluster when extreme observations

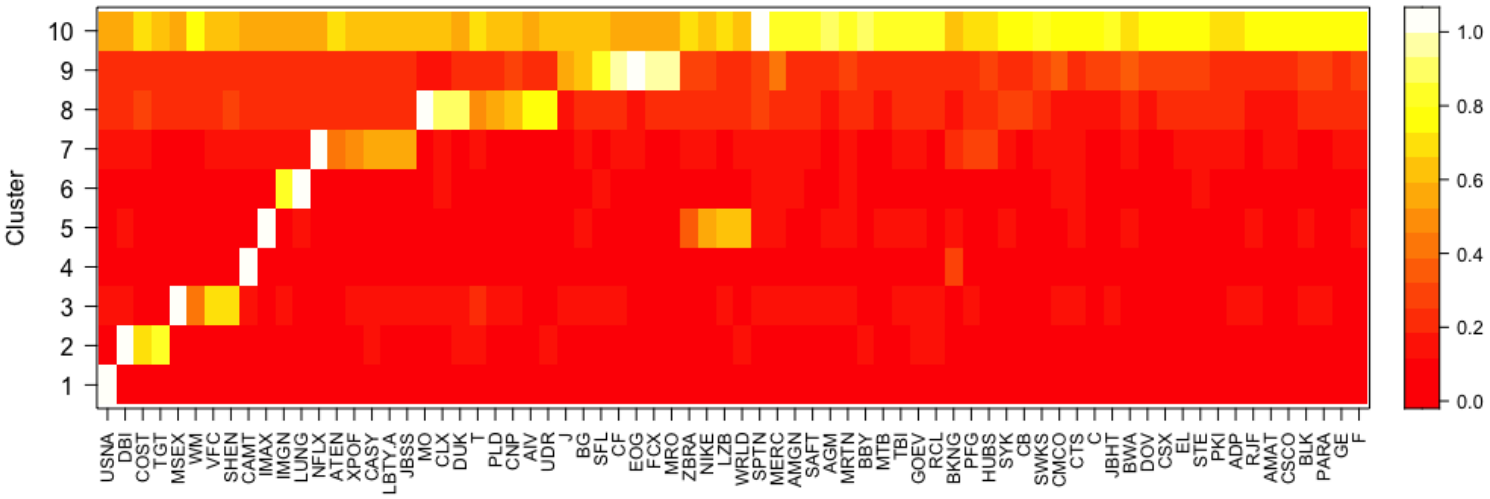


Figure 6: Result for spherical k-means clustering for k = 10. Each row corresponds to a cluster centre. The lightest colours resemble the assets with the largest value in their clusters.

Table 3: Overview of which asset belongs to which cluster for $k = 10$

Cluster	Companies
1	USNA
2	DBI, COST, TGT
3	MSEX, VFC, SHEN
4	CAMT
5	IMAX
6	IMGN, LUNG
7	NFLX
8	MO, CLX, DUK, CNP, AIV, UDR
9	SFL, CF, EOG, FCX, MRO
10	WM , ATEN, XPOF, CASY, LBTY.A, JBSS, T, PLD, J, BG, ZBRA, NKE, LZB, WRLD, SPTN, MERC, AMGN, SAFT, AGM, MRTN, BBY, MTB, TBI, GOEV, RCL, BKNG, PFG, HUBS, SYK, CB, SWKS, CMCO, CTS, C, JBHT, BWA, DOV, CSX, EL, STE, PKI, ADP, RJF, AMAT, CSCO, BLK, PARA, GE, F

Note: this table shows which asset belongs to which cluster when estimating the independence structure of extreme returns for spherical k-means for $k = 10$. Assets that are in different clusters are asymptotically independent from each other.

occur. Therefore, we also execute the spherical k-means cluster approach for $k = 24$ and see whether Cluster 10 can be divided in more clusters such that an underlying structure of the assets becomes more apparent.

Figure 7 and Table 4 show which assets belong to which cluster for $k = 24$. Most clusters contain only one asset. This means that these individual assets are asymptotically independent from each other and other clusters containing more assets. Some underlying structure is uncovered for the largest cluster of the previous estimation, however, Cluster 24 is still quite large with a cardinality of 37. So, around half of all assets are asymptotically dependent on each other when extreme observations occur. Furthermore, some clusters can be identified to represent an industry group, suggesting that assets within an industry group are dependent on each other. Cluster 17 represents the real estate industry group. Furthermore, two out of three assets in Cluster 18 are from the utilities sector. Finally, there is again an energy and materials cluster discernible, this time as Cluster 23.

The asymptotic independence between the clusters can be translated to the estimated graphs in Section 5.2 in the sense that assets that are not in the same cluster should not be linked with an

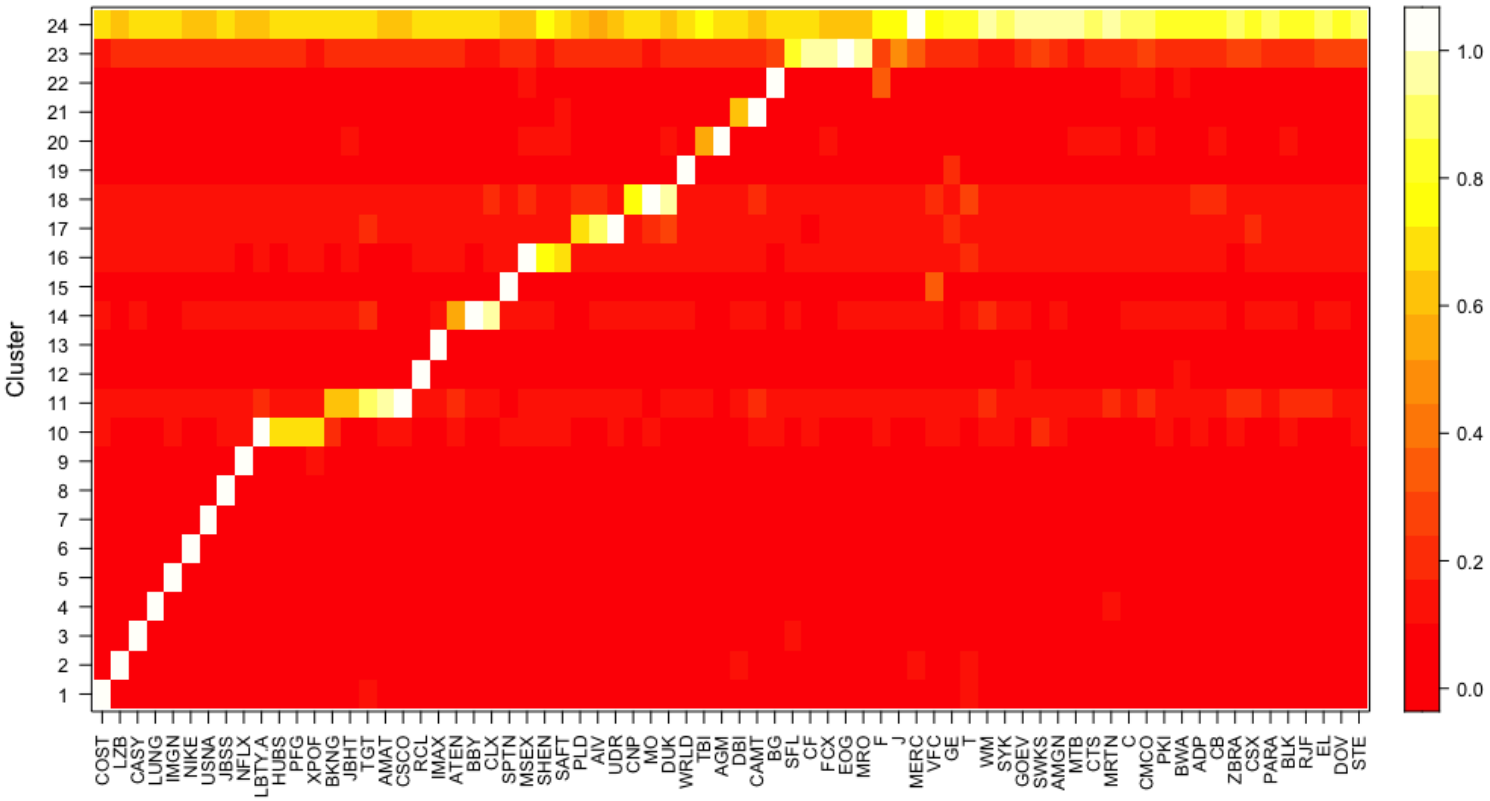


Figure 7: Result for spherical k-means clustering for k = 24. Each row corresponds to a cluster centre. The lightest colours resemble the assets with the largest value in their clusters.

Table 4: Overview of which asset belongs to which cluster for $k = 24$

Cluster	Companies
1	COST
2	LZB
3	CASY
4	LUNG
5	IMGN
6	NKE
7	USNA
8	JBSS
9	NFLX
10	LBTY.A
11	TGT, AMAT, CSCO
12	RCL
13	IMAX
14	BBY, CLX
15	SPTN
16	MSEX, SHEN
17	PLD, AIV, UDR
18	CNP, MO, DUK
19	WRLD
20	AGM
21	CAMT
22	BG
23	SFL, CF, FCX, EOG, MRO
24	HUBS, PFG, XPOF, BKNG, JBHT, ATEN, SAFT, TBI, DBI, F, J, MERC, VFC, GE, T, WM, SYK, GOEV, SWKS, AMGN, MTB, CTS, MRTN, C, CMCO, PKI, BWA, ADP, CB, ZBRA, CSX, PARA, BLK, RJF, EL, DOV, STE

Note: this table shows which asset belongs to which cluster when estimating the independence structure of extreme returns for spherical k-means for $k = 24$. Assets that are in different clusters are asymptotically independent from each other.

edge. However, this cannot hold for all assets since a graph would then disconnect and consist of separate cliques only and that is why an assumption of the graphical model is that not all assets are asymptotically independent. In that sense, the spherical k-means approach is more resolute in claiming dependence between assets.

The independence structures as a result of graphical lasso and spherical k-means overlap to a

certain extent. We would expect that the peripheral assets in the graph are similar to the assets in the smaller clusters. Yet, only five assets that are in the periphery of the graph are contained in clusters other than Cluster 10 when estimating for 10 clusters. This increases when the number of clusters increases, namely there are 14 assets in the peripheral portfolio that are not in the largest cluster when estimating spherical k-means clustering for $k = 24$.

5.4 Portfolio diversification

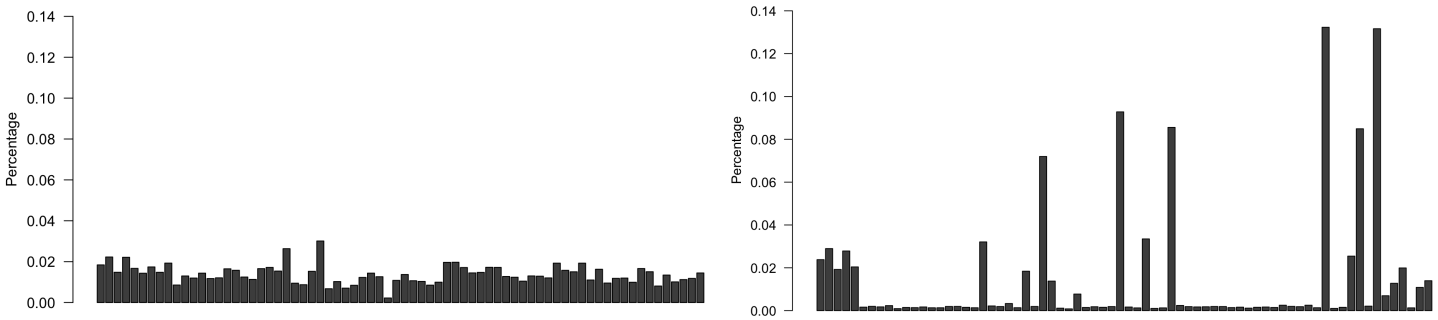
Now that we know what the portfolios are, it is possible to measure how diverse they are. In the previous sections, diversification is enhanced by selecting assets that are conditionally or asymptotically independent from each other when extreme values occur. In addition, we assumed equally weighted portfolios, which also strengthens the diversification of a portfolio. Now, the diversification of the portfolios is investigated explicitly by measuring the diversification ratio (DR) and the standard deviation of the marginal risk contribution (MRC) which were introduced in Section 4.6.1.

Table 5: Diversification measures of the portfolios

	Central	**Peripheral**	**Skmeans 10**	**Skmeans 24**	**Equally weighted**	**Maximum diversification**
DR	1.637	1.573	1.794	1.754	1.632	1.653
Std. MRC	0.017	0.013	0.028	0.019	0.004	0.012

Note: here the diversification ratio and standard deviation of the marginal risk contribution of the assets are given. These measures are estimated using daily out-of-sample returns of the assets that are contained in the various portfolios.

Table 5 presents the results of these measures. For the DR, the spherical k-means portfolios give the most diversified result. So, the increase in diversification as a result of combining the assets in a portfolio is largest for these portfolios. In addition, the peripheral portfolio shows lowest DR. This may be due to the volatility of the individual assets that are contained in this portfolio being rather low themselves. The MRC measures that risk is most equally distributed for the benchmark portfolios. Thereafter, the peripheral portfolio has lowest standard deviation of MRC. To illustrate this measure in more detail, Figure 8 gives the relative risk of each asset in percentages for the most and least diverse portfolios, the equally weighted and spherical k-means for $k = 10$ portfolios, respectively.



(a) Equally weighted portfolio

(b) Spherical 10-means portfolio

Figure 8: Marginal risk contribution in percentages of the assets in the equally weighted portfolio of all 72 assets and spherical 10-means portfolio

The figure above shows that the risk contributions of each asset in the equally weighted portfolio are almost all in the range of 0%-4% which is as expected with an equally weighted portfolio of all assets as the weights in each asset is approximately 1.4%. On the other hand, Figure 8b shows that one asset makes up for almost 15% of the risk of the whole portfolio whereas another contributes less than 1%. This is the result of the weights of the assets not being equal, assets with higher weights will have a relatively higher contribution to the total risk of the portfolio. A high volatility of the MRC is not desirable for diversification because the portfolio's risk may be impacted considerably by a shift in only one or some of the asset returns.

5.5 Portfolio performance

In addition to the portfolios' diversities, it is also interesting to review how they perform. To evaluate their performance, the Sharpe ratio for the out-of-sample period is calculated. The portfolio returns are computed as the asset returns of the portfolio times the weights of that portfolio. Below in Table 6, the results for the different portfolios are shown.

Taken generally, the portfolios with high excess portfolio return also show relatively high variance and vice versa, so expected higher returns are associated with higher risk. Except for the maximum diversification portfolio, which has highest average excess return and lowest portfolio volatility. The runner up regarding Sharpe ratio is the portfolio of peripheral assets. This portfolio holds a lot of small market capitalisation assets and these assets often show higher returns than larger firms according to the asset pricing anomaly. To take these numbers into perspective, a good annual Sharpe ratio is around 1, so the portfolios are not considered very good investments. In addition, we test whether the Sharpe ratio of, for example, the maximum

Table 6: Performance for the different portfolios

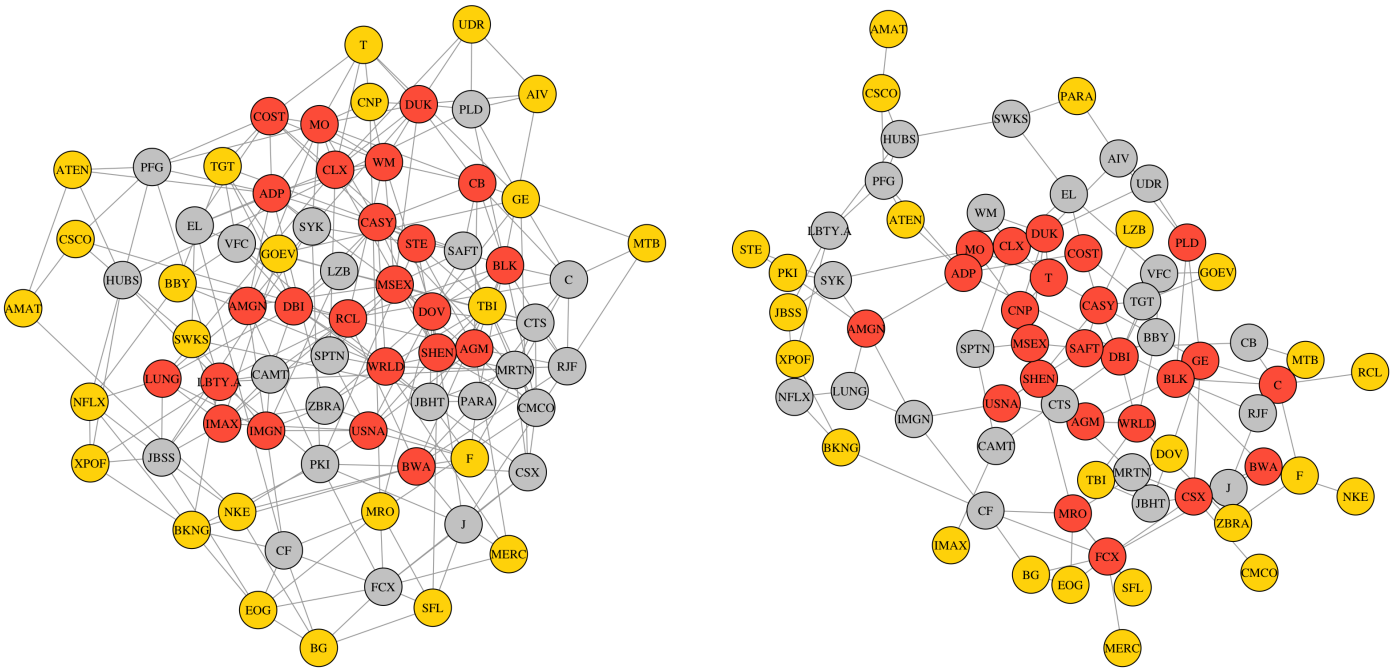
	Central	**Peripheral**	**Skmeans 10**	**Skmeans 24**	**Equally weighted**	**Maximum diversification**
μ_p^{ex}	0.024	0.043	0.031	0.038	0.036	0.047
σ_p	1.533	1.558	1.582	1.469	1.476	1.458
SR	0.244	0.435	0.308	0.413	0.387	0.511

Note: the table shows the daily mean of the excess returns and the standard deviation of gross returns. In addition, this table presents the annual Sharpe ratio for the out-of-sample data of each portfolio. The mean of the excess returns is in percentages, so the central asset has an average daily excess return of 0.024%.

diversification portfolio is significantly higher than the ratio of the central assets portfolio. It is found that non of the portfolios significantly outperform another in terms of Sharpe ratio as the p-values of all tests are all rather large. The exact results can be found in Appendix E.

5.6 Robustness of graphical model

In the previous, the portfolio selection method is executed for the graph with optimal ρ^*. However, this graph is optimal given that the model assumption is correct. Therefore, we review other values for ρ to find whether the central and peripheral assets are still the same for more or less dense graphs. In addition, the performance of these graphs is evaluated and compared to the graph for optimal ρ^*. Below in Figure 9, the estimated graphs for $\rho = 0.045$ and $\rho = 0.057$ are shown. Again the yellow vertices represent the peripheral assets, the central assets are red and the assets that do not belong to either of these sets are labelled grey. The graph for $\rho = 0.045$ is denser with 261 edges, whereas the graph for $\rho = 0.057$ only has 138 edges.



(a) Graph for $\rho = 0.045$

(b) Graph for $\rho = 0.057$

Figure 9: Hüsler-Reiss graphical model for different values for ρ

The resulting central and peripheral portfolios for the different graphs overlap considerably with the graph found with the optimal ρ. In Table 7, there is an overview of the assets that belong to each category. The assets are ordered in such a way that their overlap is more clear and not ranked as in previous tables. All central portfolios hold the first 13 assets on the left side of this table. In addition, the central assets as a result of the regularisation parameter equal to 0.045 and 0.057 have two and five additional overlapping assets respectively. For the peripheral portfolio the first 12 assets overlap for the three graphs. Moreover, the portfolio with optimal ρ^* has 15 assets in common with the graph estimated with $\rho = 0.045$ and 18 mutual assets with the model estimated with $\rho = 0.057$.

Yet, moving from a dense to a more sparse graph can also mean that central assets become peripheral because these assets lose connections in sparser graphs. This is the case for assets LUNG, RCL and STE which are in the central portfolio for $\rho = 0.045$ and in the peripheral portfolio for optimal ρ^*. In contrast, the relative centrality of an asset can increase when a graph becomes sparser because other assets become less central. This is true for the asset T, which is in the periphery for the lower value for ρ and is contained in the central portfolio for optimal ρ^* and $\rho = 0.057$

Moving on to the performance of the portfolios that result from the less and more sparse graphs,

Table 7: Overview of assets in the central and peripheral portfolio for various graphs based on different values for regularisation parameter ρ

	Central portfolio			**Peripheral portfolio**		
	$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$	$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$
1	ADP	ADP	ADP	AMAT	AMAT	AMAT
2	AGM	AGM	AGM	ATEN	ATEN	ATEN
3	BLK	BLK	BLK	BG	BG	BG
4	BWA	BWA	BWA	CSCO	CSCO	CSCO
5	CASY	CASY	CASY	EOG	EOG	EOG
6	CLX	CLX	CLX	F	F	F
7	DBI	DBI	DBI	GOEV	GOEV	GOEV
8	DUK	DUK	DUK	MERC	MERC	MERC
9	MO	MO	MO	MTB	MTB	MTB
10	MSEX	MSEX	MSEX	NKE	NKE	NKE
11	SHEN	SHEN	SHEN	TBI	TBI	TBI
12	USNA	USNA	USNA	XPOF	XPOF	XPOF
13	WRLD	WRLD	WRLD	BBY	BBY	BKNG
14	IMGN	IMGN	AMGN	NFLX	NFLX	DOV
15	LBTY.A	LBTY.A	CNP	UDR	UDR	IMAX
16	C	AMGN	C	CMCO	BKNG	CMCO
17	CSX	CB	CSX	JBSS	SWKS	JBSS
18	FCX	COST	FCX	LZB	MRO	LZB
19	SAFT	DOV	SAFT	PKI	GE	PKI
20	T	IMAX	T	RCL	T	RCL
21	CF	LUNG	COST	STE	TGT	STE
22	CTS	RCL	GE	CAMT	AIV	PARA
23	EL	STE	MRO	LUNG	CNP	SFL
24	RJF	WM	PLD	SPTN	SFL	ZBRA

Note: this table shows the assets that are contained in the central and peripheral portfolios for values of the regularisation parameter equal to $\rho = 0.045$ and $\rho = 0.057$ in the graphical lasso estimate in addition to the theoretically optimal value for ρ^*. Assets are not ranked as in previous tables, to highlight the overlap the portfolios have.

the daily excess mean and standard deviation together with the yearly Sharpe ratio of the three central and three peripheral portfolios are given in Table 8. This table shows that the Sharpe ratios of the portfolios for various values of the regularisation parameter differ somewhat, yet there are no large differences. For the sake of completeness, it is tested whether the Sharpe ratios are significantly different with the test from Section 4.6.2. We find that there is no portfolio

Table 8: Performance of the central and peripheral portfolios that result from less and more dense graphs

	Central			**Peripheral**		
	$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$	$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$
μ_p^{ex}	0.024	0.027	0.024	0.043	0.027	0.032
σ_p	1.533	1.434	1.479	1.558	1.543	1.653
SR	0.244	0.301	0.261	0.435	0.281	0.309

Note: the table shows the daily mean of the excess returns and the standard deviation of gross returns. In addition, this table presents the annual Sharpe ratio for the out-of-sample data of the central and peripheral portfolios with various values for the regularisation parameter. The mean of the excess returns is in percentages, so the central asset has an average daily excess return of 0.024%.

that significantly outperforms the other based on Sharpe ratio. Thus, when the regularisation parameter is slightly different, this does not have a large impact on the performance of the resulting portfolios. For the exact results, see Appendix E.

6 Conclusion

To summarise, in this paper, two data-driven methods to evaluate the dependence structures for extreme asset losses have been examined. Firstly, we have investigated a graphical model based on ℓ_1-regularisation and found the optimal graphical dependence between the assets. Secondly, the independence structure of the assets was found with the spherical k-means approach. The methods were the foundation for the asset selection to build a well-diversified portfolio.

We have found that the underlying independence structures of both approaches are related to the GICS industry (group) categorisation to some extent. Assets from the same industry (group) were found to be closely connected in the graphical model and they were also observed to form clusters. In addition, the assets that were estimated to have low conditional dependence or those that are asymptotically independent of each other as a result of graphical lasso and spherical k-means overlap moderately. Namely, of the peripheral assets, five assets formed one-asset-clusters for the spherical 10-means estimation, for the spherical 24-means estimation this number is 14.

Furthermore, we have found that the distribution of marginal risk contribution of the peripheral portfolio is rather good compared to the maximum diversification portfolio, which is constructed to optimise the diversification measures. On the other hand, although the spherical k-means portfolios have high diversification benefits from combining the assets in a portfolio, the risk

coming from the assets is not very equally distributed. This is due to the fact that these portfolios do not hold equal weights among the assets it holds.

Regarding out-of-sample performance, the portfolios do not have a very high Sharpe ratio. Among the data-driven portfolios, the peripheral portfolio shows the best Sharpe ratio, however it is not statistically different from the other ratios. So, we cannot say that it outperforms the other portfolios concerning this measure.

In addition, when the graphical model is not evaluated with the optimal value for the regularisation parameter, it has been shown that the central and peripheral portfolios of more and less dense graphs still hold the majority of assets that is held by the portfolios that follow from the graph with optimal ρ. The changes that do result from the other estimates do not influence the performance of the portfolios significantly. Therefore, we conclude that asset selection based on graphical models is robust regarding the specification of the regularisation parameter.

The research is limited regarding two aspects. Firstly, the data consists only of 72 companies, whereas a real-world investor would have a much wider range of assets to invest in to diversify their portfolio, including alternative investment opportunities than stocks. However, including many more will make the empirical analysis more time consuming. In addition, all companies are US originated and therefore may have higher dependencies, moreover no foreign assets were used to diversify. The methods that are used may be susceptive to changes in the data, so switching some companies with new data will result in different graphical structures and clusters and this may produce different asset selection decisions.

Secondly, the asset selection procedures are based on extreme values, however for the portfolio evaluation, also non-extreme data is included since real-life data does not only exist of extreme data. The reason for doing it in this way, is that investors do not know when a crisis will occur and thus they cannot switch beforehand to a portfolio that is resilient in turbulent times. A way to encompass this may be to use an extreme observations indicator that signals when to use a strategy for extremes and when to work with another investment strategy that works well during non-extreme times.

Topics of further research may include tackling these challenges. For example, one may execute a robustness analysis on the assets that are included in the data set and how the portfolio changes when certain assets are left out of the data set. Moreover, a direction of further research may be to find more dependency structure in the large clusters that remain after spherical k-means clustering, such that the weights in this large cluster can be sustained on something else than equal weights. For example, by only applying this method to the assets that are in this cluster.

Now the cardinality of the large clusters implies that a lot of assets are driven by the same risk factor when extremes occur. Another way to find more structure in these clusters is thus to find these risk factors and correct for them in the returns filtering process.

References

Babu, M. S., Geethanjali, N., and Satyanarayana, B. (2012). Clustering approach to stock market prediction. *International Journal of Advanced Networking and Applications*, 3(4):1281.

Bessembinder, H. (2018). Do stocks outperform treasury bills? *Journal of financial economics*, 129(3):440–457.

Choueifaty, Y. and Coignard, Y. (2008). Toward maximum diversification. *The Journal of Portfolio Management*, 35(1):40–51.

Clemente, G. P., Grassi, R., and Hitaj, A. (2021). Asset allocation: new evidence through network approaches. *Annals of Operations Research*, 299(1):61–80.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The review of Financial studies*, 22(5):1915–1953.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932.

Engelke, S. and Ivanovs, J. (2021). Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8:241–270.

Engelke, S. and Volgushev, S. (2020). Structure learning for extremal tree models. *arXiv preprint arXiv:2012.06179*.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fomichov, V. and Ivanovs, J. (2022). Spherical clustering in detection of groups of concomitant extremes. *Biometrika*.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Green, R. C. and Hollifield, B. (1992). When will mean-variance efficient portfolios be well diversified? *The Journal of Finance*, 47(5):1785–1809.

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

Jagannathan, R. and Ma, T. (2003). Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683.

Janßen, A. and Wan, P. (2020). k-means clustering of extremes. *Electronic Journal of Statistics*, 14(1):1211–1233.

Ledoit, O. and Wolf, M. (2004). Honey, i shrunk the sample covariance matrix. *The Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060.

León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., and Niño, J. (2017). Clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, 108:1334–1343.

Maillard, S., Roncalli, T., and Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management*, 36(4):60–70.

Mainik, G., Mitov, G., and Rüschendorf, L. (2015). Portfolio optimization for heavy-tailed assets: Extreme risk index vs. markowitz. *Journal of Empirical Finance*, 32:115–134.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.

Markowitz, H. M. (1952). Portfolio selection, 1959. *Journal of Finance*, 7:7791.

Mazumder, R. and Hastie, T. (2012). The graphical lasso: New insights and alternatives. *Electronic journal of statistics*, 6:21–25.

Michaud, R. O. (1989). The markowitz optimization enigma: Is 'optimized'optimal? *Financial analysts journal*, 45(1):31–42.

Onnela, J.-P., Chakraborti, A., Kaski, K., and Kertiész, J. (2002). Dynamic asset trees and portfolio analysis. *The European Physical Journal B-Condensed Matter and Complex Systems*, 30(3):285–288.

Peralta, G. and Zareei, A. (2016). A network approach to portfolio selection. *Journal of Empirical Finance*, 38:157–180.

Pozzi, F., Di Matteo, T., and Aste, T. (2013). Spread of risk across financial markets: better to invest in the peripheries. *Scientific reports*, 3(1):1–7.

Resnick, S. I. (2007). *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Science & Business Media.

Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, 28(1):82–108.

Zaheer, K., Aslam, F., Mohmand, Y. T., and Ferreira, P. (2022). Temporal changes in global stock markets during covid-19: an analysis of dynamic networks. *China Finance Review International*.

A Company information

Table 9: Overview of industries, industry groups and companies used for estimation together with their market capitalisation and stock symbol (ticker)

Sector	Industry group	Company	Market cap ($B)	Ticker
Energy	Energy	EOG Resources, Inc.	72.3	EOG
		Marathon Oil Corporation	19.1	MRO
		SFL Corporation Ltd.	1.3	SFL
Materials	Materials	Freeport-McMoRan Inc.	71.9	FCX
		CF Industries Holdings, Inc.	22.8	CF
		Mercer International Inc.	1.1	MERC
Industrials	Capital Goods	General Electric Company	100.0	GE
		Dover Corporation	20.8	DOV
		Columbus McKinnon Corporation	1.1	CMCO
	Commercial & Professional Services	Waste Management	66.3	WM
		Jacobs Engineering Group Inc.	18.5	J
		TrueBlue, Inc.	1.1	TBI
	Transport	CSX Corporation	76.0	CSX
		J.B. Hunt Transport Services, Inc.	18.0	JBHT
		Marten Transport, Ltd.	1.4	MRTN
Consumer Discretionary	Automobiles & Components	Ford Motor Company	62.3	F
		BorgWarner Inc.	9.3	BWA
		Canoo Inc.	1.2	GOEV
	Consumer Durables & Apparel	Nike, Inc.	174.2	NKE
		V.F. Corporation	22.0	VFC
		La-Z-Boy Incorporated	1.2	LZB
	Consumer Services	Booking Holdings Inc.	91.6	BKNG
		Royal Caribbean Cruises Ltd.	20.9	RCL
		Xponential Fitness, Inc	1.1	XPOF
	Retailing	Target Corporation	108.1	TGT
		Best Buy Co.	21.4	BBY
		Designer Brands Inc.	1.1	DBI
Consumer Staples	Food & Staples Retailing	Costco Wholesale Corporation	201.5	COST
		Casey's General Stores, Inc.	7.9	CASY
		SpartanNash Company	1.2	SPTN
	Food, Beverages & Tabacco	Altria Group, Inc.	99.3	MO
		Bunge Limited	17.3	BG
		John B. Sanfilippo & Son, Inc.	1.0	JBSS
	Household & Personal Products	The Estée Lauder Companies Inc.	95.6	EL

		The Clorox Company	17.9	CLX
		e.l.f. Beauty, Inc.	1.3	ELF
Health Care	Health Care Equipment & Services	Stryker Corporation	102.0	SYK
		STERIS plc	25.0	STE
		Pulmonx Corporation	1.0	LUNG
	Pharmaceuticals, Biotechnology & Life Sciences	Amgen	136.4	AMGN
		PerkinElmer, Inc.	20.1	PKI
		ImmunoGen, Inc.	1.1	IMGN
Financials	Banks	Citigroup Inc.	98.9	C
		M&T Bank Corporation	20.6	MTB
		Federal Agricultural Mortgage Corporation	1.2	AGM
	Diversified Financials	BlackRock, Inc.	108.6	BLK
		Raymond James Financial, Inc.	22.8	RJF
		World Acceptance Corporation	1.2	WRLD
	Insurance	Chubb Limited	90.2	CB
		Principal Financial Group, Inc.	19.3	PFG
		Safety Insurance Group, Inc.	1.3	SAFT
Information Technology	Software & Services	Automatic Data Processing	96.7	ADP
		HubSpot, Inc.	22.1	HUBS
		A10 Networks, Inc.	1.0	ATEN
	Technology Hardware & Equipment	Cisco Systems, Inc.	216.6	CSCO
		Zebra Technologies Corporation	20.1	ZBRA
		CTS Corporation	1.1	CTS
	Semiconducters & Semiconductor Equipment	Applied Materials, Inc.	103.2	AMAT
		Skyworks Solutions, Inc.	19.8	SWKS
		Camtek Ltd.	1.3	CAMT
Communication Services	Telecommunication Services	AT&T	139.1	T
		Liberty Global plc	13.5	LBTY.A
		Shenandoah Telecommunications	1.1	SHEN
	Media & Entertainment	Netflix, Inc.	96.9	NFLX
		Paramount Global	21.1	PARA
		IMAX Corporation	1.0	IMAX
Utilities	Utilities	Duke Energy Corporation	88.6	DUK
		CenterPoint Energy, Inc.	20.3	CNP
		Middlesex Water Company	1.7	MSEX
Real Estate	Real Estate	Prologis, Inc.	124.2	PLD
		UDR, Inc.	19.6	UDR

	Apartment Investment and Management Company	1.0	AIV

Note: Overview of the 72 American companies which asset returns are used for the empirical analysis. They are selected such that there are three companies from each industry group. One with small (around 1$B), one with medium (around 20$B) and one with large (around 100$B) market capitalisation. The last column shows the ticker symbols, these abbreviation are used throughout the paper to refer to the companies.

B Assets in the maximum diversification portfolio

Table 10: Assets that are in the maximum diversification portfolio

ADP	CAMT	GE	JBSS	NFLX	STE
AMGN	COST	HUBS	LUNG	PFG	SWKS
BBY	CSX	IMAX	MRO	PKI	USNA
BLK	EL	IMGN	MTB	SPTN	XPOF

Note: The assets of the maximum diversification portfolio are found through finding the portfolio that performs best regarding the diversification ratio and the volatility of marginal risk distribution among 100,000 unique random combinations of 24 assets.

C ARMA-GARCH result

Table 11: Overview of fit for various (p,q)ARMA-(r,s)GARCH models based on the total AIC, the best fit is in bold

p	q	r	s	AIC
0	0	1	1	287.867
0	1	1	1	287.760
1	0	1	1	287.766
1	1	1	1	287.725
0	0	1	2	287.852
0	1	1	2	287.748
1	0	1	2	287.753
1	**1**	**1**	**2**	**287.714**
0	0	2	1	287.915
0	1	2	1	287.808
1	0	2	1	287.814
1	1	2	1	287.775
0	0	2	2	287.893
0	1	2	2	287.787
1	0	2	2	287.792
1	1	2	2	287.754

Note: in order to transform the in-sample asset returns into approximately i.i.d. returns, an (p,q)ARMA-(r,s)GARCH model is used. To find the best fit, various models with different values for the parameters are tested and evaluated regarding their AIC score.

D Centrality measures

Table 12: Separate results of the closeness and degree centrality measures of the graphical model for optimal ρ

Rank	Closeness	Degree	Rank	Closeness	Degree
1 (most central)	MO	MO	49	HUBS	BBY
2	SHEN	DUK	50	F	MRTN
3	AGM	SHEN	51	IMAX	DOV
4	CASY	RJF	52	MERC	SFL
5	BLK	BLK	53	LUNG	EOG
6	IMGN	C	54	JBSS	UDR
7	CLX	FCX	55	AIV	CAMT
8	RJF	LBTY.A	56	PLD	CSCO
9	ADP	ADP	57	MTB	MTB
10	USNA	IMGN	58	CAMT	PKI
11	CF	CLX	59	VFC	LUNG
12	EL	CASY	60	BBY	STE
13	DUK	DBI	61	EOG	BG
14	CTS	BWA	62	BG	SPTN
15	SWKS	CSX	63	STE	XPOF
16	MSEX	CF	64	NFLX	RCL
17	LBTY.A	MSEX	65	ATEN	NKE
18	C	PFG	66	UDR	GOEV
19	SAFT	WRLD	67	XPOF	F
20	COST	AGM	68	TBI	CMCO
21	BWA	TGT	69	GOEV	AMAT
22	CB	JBHT	70	PKI	LZB
23	DBI	MRO	71	CSCO	TBI
24	T	T	72 (least central)	AMAT	MERC

Note: to select the assets that belong to the peripheral and central portfolio, the assets are ranked according to their closeness and degree centrality. This table shows the separate rankings of the two measures. On the left side of the table, the assets are shown that are most central according to the measures. On the right side the least central assets are ranked according to their closeness and degree centrality.

E Significance tests of the portfolios' Sharpe ratios

Table 13: p-values of the significance tests for portfolios' Sharpe ratios

	Central	**Peripheral**	**Skmeans 10**	**Skmeans 24**	**Equally weighted**	**Maximum diversification**
Central	-					
Peripheral	0.301	-				
Skmeans 10	0.733	0.483	-			
Skmeans 24	0.252	0.864	0.395	-		
Equally weighted	0.209	0.619	0.622	0.807	-	
Maximum diversification	0.195	0.607	0.179	0.477	0.401	-

Note: this table represents the p-values of the test that find whether the out-of-sample Sharpe ratios of the portfolios are significantly different from each other. The p-values are found using HAC inference with Parzen kernel.

Table 14: p-values of the significance tests for Sharpe ratios of the portfolios resulting from the robustness analysis

		Central			**Peripheral**		
		$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$	$\rho^* = 0.053$	$\rho = 0.045$	$\rho = 0.057$
Central	$\rho^* = 0.053$	-					
	$\rho = 0.045$	0.301	-				
	$\rho = 0.057$	0.733	0.483	-			
Peripheral	$\rho^* = 0.053$	0.252	0.864	0.395	-		
	$\rho = 0.045$	0.209	0.619	0.622	0.807	-	
	$\rho = 0.057$	0.195	0.607	0.179	0.477	0.401	-

Note: this table represents the p-values of the test that find whether the out-of-sample Sharpe ratios of the central and peripheral portfolios for different values of ρ are significantly different from each other. The p-values are found using HAC inference with Parzen kernel.