UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

ELECTRONICS AND COMPUTER SCIENCE

The Inference and Analysis of Correlation and Partial Correlation Financial Networks

Author:
Tristan Millington

Supervisor:
Professor Mahesan NIRANJAN

A report submitted in fulfillment of the requirements for the PhD in the Vision, Learning and Control
School of Electronics and Computer Science

May 14, 2021

Declaration of Authorship

I, Tristan Millington, declare that this thesis titled, "The Inference and Analysis of Correlation and Partial Correlation Financial Networks" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FPSE
School of Electronics and Computer Science

PhD

The Inference and Analysis of Correlation and Partial Correlation Financial Networks

by Tristan Millington

To understand risk in a financial market we must understand how asset prices are related. By using correlation measures we can quantify the relationships between asset pairs, and then using network science we can then attempt to study the system as a whole. In this thesis we explore the use of various correlation and partial correlation estimators to estimate a financial network from returns data. We then show how these networks differ from the standard Pearson correlation models and attempt to evaluate their use.

We firstly explore the use of sparse precision matrix estimators, and compare their success in selecting a known underlying model to simply thresholding the sample covariance matrix. Surprisingly we find that thresholding the sample covariance matrix is competitive with these sparse precision matrix estimators. Next we look at using a selection of these estimators for portfolio optimization. We find that in general they do not outperform non-sparse methods, such as the Ledoit-Wolf shrinkage estimators, but can provide some added robustness, and do force diversification upon the portfolios.

We then look at constructing networks using these precision matrix estimations. Firstly we use the Ledoit-Wolf shrinkage method to construct partial correlation networks of the S&P500. These partial correlation networks have a significantly less variable largest eigenvalue than a correlation network, indicating the effect of the market has been removed, but in fact are more unstable than their correlation counterparts, with both the largest eigenvector and community structure changing significantly more between adjacent time periods. Furthermore we have less success in uncovering the underlying sector structure in these partial correlation networks compared to the equivalent correlation ones.

These Ledoit-Wolf estimated networks are dense, which can inhibit interpretability. Therefore next we look at using a sparse precision matrix estimator, the SPACE method. Again these networks seem quite unstable, with a large number of edge changes between networks adjacent in time, indicating that partial correlation networks in general are more unstable than their correlation counterparts.

Next we explore the use of rank correlation methods for the construction of minimum spanning trees from financial returns, and explore how these compare to those constructed using Pearson correlation. We find that the trees constructed using these rank methods correlation tend to be more stable and maintain more edges over the dataset than those constructed using Pearson correlation and the trees have similar topologies. We also explore how deviations from Gaussianity drive differences in the trees. There is little correlation between MST differences and deviations from univariate Gaussianity, but if we use quantile normalization to force

the dataset to be univariate Gaussian then the differences between the MSTs drops, indicating this does have an effect.

Finally we look at how the similarity and stability of correlation networks changes during times of market calm and market stress. Using some simple measures, such as the change and standard deviation of the entries in the leading eigenvector and the mean L_2 difference between nodes, we look at three different markets, the US, UK and Germany, and find that the UK and US markets become more similar and more stable during times of market stress, but the German market does not see such effects.

Contents

List of Figures

List of Tables

For Kieran

Chapter 1

Introduction

1.1 Introduction and Motivation

Financial markets are a complex, dynamic system. Many different investors, with a wide variety of objectives, interact together to agree on prices for assets. Regulators also wish to observe, and occasionally intervene in markets, to ensure their stability.

A range of factors contribute to the price of an asset, including the price of a commodity, interest rates, past performance of the asset and government policy. Generally investors are interested in maximising the return and minimizing the risk of their portfolios (the set of assets they own). The return of a portfolio is how much the price of the assets it contains rises or falls during a period of time, while the risk of a portfolio is a measure of the chance of the price changing significantly. In general financial practitioners are interested in two forms of risk, that from the perspective of a portfolio, and that from the perspective of the system as a whole (known as *systematic risk*)

Investors are concerned with the former. Due to a desire to reduce risk they own portfolios made up of multiple assets. This process is referred to as diversification. To accurately assess risk in these portfolios we must understand the dynamics of the relationships between said assets. The goal here is also often to minimize the amount of risk inherent in the portfolio. Furthermore they may also be interested in patterns of behaviour in the market. While "past performance is not an indicator of future results" is an important phrase to remember when it comes to investments, markets do fall into certain states, and by studying the properties of certain states, and which they transition into, we can make recommendations for investors.

Both investors and regulators tend to be interested in the latter. Unlike risk from the perspective of a portfolio, systematic risk cannot be minimized by an investor. In an ideal world, if the level of systematic risk is high, a regulator would step in. High levels of systematic risk can lead to market crashes, and potentially even a breakdown of the financial system. Therefore it is desirable to quantify the amount of systematic risk in order to inform a regulator when it might be wise to step in, or for an individual investor to remove themselves from the market for a period of time.

Network science can provide a toolkit to allow us to study the markets, and to understand the relationships that occur in them. By using both global and node level measures we can quantify how the entire system and individual components behave and respond, and this helps us to understand risk from both perspectives.

In the context of financial networks, we are not given the true underlying network and so must infer it from data. The usual choice is to use the Pearson correlation between asset returns. This has the advantages of being simple the interpret and calculate. The challenge is that asset returns are not stationary, so a common solution is to take a window of data where the returns can be assumed to be approximately

stationary. This necessitates a trade off between window size and stationarity. A small window means this assumption of stationarity is more accurate, but this can lead to the challenge of having more dimensions than samples. Financial returns are also driven by the overall market as well as factors that influence individual companies, meaning that there are spurious or indirect correlations which a user may desire to have removed. The problems mentioned above are found in other fields (e.g. biology, neuroscience) where the acquisition of extra samples can be expensive. Therefore in the first few chapters we see if tools from these fields can be used in a financial context. In particular we focus on sparse precision matrix estimation, where the methods are designed to reliably infer precision matrices in $p >> n$ situations. Precision matrices can also be used to construct a partial correlation matrix, which removes these indirect correlations. Chapters 3, 4 and 6 focus on this. The results in these situations seem to indicate that the networks are very unstable, but this could be due to the instability of the lasso on highly correlated data. Therefore to rectify this situation in Chapter 5 we look at the inference of precision matrices and then partial correlation networks by using a non-sparse estimator, the Ledoit-Wolf shrinkage methods.

Pearson correlation is also not a robust estimator, and so can be badly affected by outliers and non-linear relationships. Both of these are often present in financial returns, and so in Chapter 7 we explore the use of rank correlation methods for construction of financial networks, as rank correlation is not susceptible to either of these issues.

As previously mentioned, one of the goals of constructing and analyzing these networks is to help inform how an investor might diversify their portfolios. Therefore, in chapter 8 we investigate some of the properties of Pearson correlation networks that relate to the goals of diversification vary during times of market calm and stress. Below we give a more detailed description of each chapter.

1.2 Contributions

The main contributions of this thesis are

- **The application of sparse high dimensional precision matrix estimators to financial data**. We use these estimators for portfolio optimization (Chapter 4) and for inferring a network of company relationships (Chapter 6)
- **Analysis of partial correlation financial networks**. We compare and contrast correlation and partial correlation networks inferred from financial data (Chapter 5, Chapter 6, Appendix A).
- **The comparison of Pearson, Spearman and Kendall's τ correlation coefficient for the construction of minimum spanning trees** (Chapter 7). Studying the difference between trees inferred using these correlation coefficients, we show that the trees inferred using rank correlation are more stable and less affected by periods of disruption, and tend to have slightly more reliable edge selection
- **The study of correlation based financial networks in times of market stress and calm** (Chapter 8). Using some simple network measures and a community detection algorithm, we show these networks could be considered more stable during times of market stress

1.3 Structure

Chapter 2 provides a review of the literature in inference and analysis of financial networks from returns data. In it we detail the various methods designed to remove noisy from the matrices, including topological based methods (e.g. minimum spanning trees, planar maximally filtered graph) and spectral based methods (e.g. random matrix theory). We include how these networks change over time, alternative methods of inferring a financial network (e.g. correlation, partial correlation, mutual information) and a software package that implements some of the methods discussed. Some applications of these networks are shown (e.g. portfolio selection) and finally, how they may be used as inputs into clustering algorithms, which can then be used for portfolio diversification.

Chapter 3 provides a brief review of the estimation of precision matrices, mostly focusing on imposing sparsity. We show how linear regression can be used to estimate a row of the precision matrix, and from that how various advances in regularised linear regression are then used to estimate precision matrices. We then compare the performance of a selection of precision matrix estimators on Gaussian data. This is quantified using the area under the ROC curve and the F_1 score when a regularization parameter is selected using cross validation. Interestingly we find that thresholding the sample covariance matrix is very competitive with even the best sparse precision matrix estimation methods, and is computationally trivial.

Chapter 4 focuses on the use of sparse precision matrix estimation methods in portfolio optimization. Noting that portfolio optimization suffers from many of the problems these estimators are designed to fix - e.g. more dimensions than samples, noisy data, we apply the graphical lasso, SCIO, Scaled Lasso and Ledoit-Wolf shrinkage method to financial returns in the hope they will improve portfolio performance and robustness. We find that the Ledoit-Wolf shrinkage method is the best at reducing out-of-sample risk, but that the sparse methods can provide an improvement to the robustness of portfolios, and to the turnover and size of asset holdings.

Chapter 5 looks at the inference of partial correlation financial networks using the Ledoit-Wolf shrinkage estimators and compares them to correlation based ones. Firstly we show that the partial correlation networks have a smaller and much less variable intensity than the correlation networks, but in fact are less stable. We look at the centrality of the various sectors in the graph using degree centrality and eigenvector centrality, finding that sector centralities move together during the 2009 market crash and that the financial sector generally has a higher mean centrality over most of the dataset. Exploring the use of these centrality measures for portfolio construction, we shown there is mild correlation between the in-sample centrality and the out of sample Sharpe ratio but there is negative correlation between the in-sample centrality and out of sample risk. Finally we use a community detection method to study how the networks reflect the underlying sector structure and study how stable these communities are over time. Further analysis of partial correlation networks is done in appendix A, where we take a topological approach to comparing the correlation and partial correlation networks. Here we again find the partial correlation networks are less stable, less clustered and could be considered noisier.

Chapter 6 looks at using the SPACE algorithm, which estimates a sparse partial correlation matrix, to estimate partial correlation networks for S&P500 returns and looks at how they have evolved over time. We see that companies tend to have more connections to those in the same sector and some sectors tend to be more self contained than others. By measuring the centrality of the various sectors in the network we find that the financial sector is regarded as the most important for the

majority of the dataset. Finally we show there is mild negative correlation between the centrality of a company and its out-of-sample risk.

In Chapter 7 we look at the construction of minimum spanning trees from financial returns. This is an often used method to study relationships in the financial markets due to the simplicity of construction, interpretation, and the removal of noisy edges. However most of the work on this topic tends to use the Pearson correlation coefficient, which relies on the assumption of normality and can be brittle to the presence of outliers, neither of which is ideal for the study of financial returns. We study the inference of MSTs over a time period from daily US, UK and German financial returns using both the Pearson and two correlation coefficients, Spearman and Kendall's τ. We find that the trees constructed using rank correlation tend to change less and maintain more edges over the dataset than those constructed using Pearson correlation. Generally the trees have similar topologies, irrelevant of the coefficient used. Using a bootstrap method we show the rank correlations can have a small impact on the robustness of the MSTs, but the construction procedure has the largest effect.

In Chapter 8 we use network theory to study the correlations between stocks and how this varies over time, using daily returns from the S&P500 (US), FTSE100 (UK) and DAX30 (Germany). Our conclusion is that stocks returns tend to become more similar during times of market disruption for the US and UK markets - implying that nodes that were once dissimilar (and perhaps therefore a good choice for a low risk portfolio) are no longer so, demonstrating the difficulties of choosing a diversified portfolio. Furthermore the networks are also more stable by certain measures during these periods of disruption, contrary to expectations. However these apply less to the German market, perhaps either due to the smaller size of the German market or the structural difference of the German economy.

1.4 Publications

These publications are based purely on research conducted during my PhD

- Tristan Millington and Mahesan Niranjan. "Robust Portfolio Risk Minimization Using the Graphical Lasso." International Conference on Neural Information Processing. Springer, Cham, 2017.
- Tristan Millington and Mahesan Niranjan. "Quantifying Influence in Financial Markets via Partial Correlation Network Inference." 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, 2019.
- Tristan Millington and Mahesan Niranjan. "Partial correlation financial networks" Journal of Applied Network Science
- Tristan Millington and Mahesan Niranjan. "Construction of Minimum Spanning Trees from Financial Returns using Rank Correlation" Physica A
- Tristan Millington and Mahesan Niranjan. "Similarity and Stability in Financial Networks - How do they change during Times of Turbulence?" Physica A

Chapter 2

A Review of the Inference and Analysis of Financial Networks from Asset Returns

2.1 Introduction

The dynamics of asset returns are complex, with many different factors influencing their movement. Furthermore portfolios do not consist of a single asset, and so we are not just interested in the dynamics of one asset, but how asset prices move together. This can lead to insights in how to diversify portfolios and assess risk, both from the perspective of said portfolio and from the perspective of a market as a whole (known as systematic risk). We can use network theory to help understand these relationships, as it provides us with a set of tools that can be used to study a variety of complex systems using common measures. Networks created from a wide variety of fields show surprisingly common properties, such as community structure, small world behaviour and power law degree distributions [10] [11]. In this chapter we review work on the inference and applications of financial networks from asset returns and summarize the conclusions and insights that have been found so far.

2.2 Correlation Network Filtration

Often the genesis of studying the correlations in a financial system using a graph model is credited to Mantegna [160]. Starting with the log returns of a stock

$$r_i(t) = \log x_i(t) - \log x_i(t-1) \tag{2.1}$$

where $x_i(t)$ is a the price of asset i and time t in our dataset containing p stocks with n samples, we calculate the Pearson correlation coefficient between assets i and j as

$$c_{ij} = \frac{\sum_{i=1}^{n}(r_i(t) - \bar{r}_i)(r_j(t) - \bar{r}_j)}{\sqrt{\sum_{i=1}^{n}(r_i(t) - \bar{r}_i)^2 (r_j(t) - \bar{r}_j)^2}} \tag{2.2}$$

where $\bar{r}_i$ is the mean of r_i.

For various reasons, notably that it can be negative, correlation is not necessarily the most suitable measure to use as a distance measure. However it can be trivially turned into a ultrametric (i.e. non-negative, symmetrical, obeys the strong triangle

inequality and 0 only when $i = j$) using the following formula

$$d_{ij} = \sqrt{2(1 - c_{ij})} \tag{2.3}$$

Either the correlation or distance matrix obtained from the asset returns can be treated like an adjacency matrix, which results in a complete graph. This is difficult to interpret and visualize, and furthermore many of the edges contained will be small valued and noisy. Therefore it is desirable to remove these edges to increase interpretability and to remove noise. In this section we detail various methods used to discard these edges.

2.2.1 Minimum Spanning Tree

Mantegna [160] constructs a minimum spanning tree (MST) using the distance graph. A minimum spanning tree is a fully connected graph without any cycles and with the minimum possible edge weight. The tree can be created using Kruskal's algorithm, which proceeds as follows

1. Initialise the tree as a disconnected graph made up of the nodes in the distance graph
2. Sort edges of distance graph in ascending order and place in a list
3. For each edge between i and j in this list
   (a) If i and j are not in the same component in the MST, add an edge between them with the weight from the distance matrix

The MST model is quite restrictive but does remove noisy low valued edges from the correlation matrix. In these trees the branches tend to contain assets that are similar - for instance when inferred from stock returns then they tend to contain companies in the same sector. Effectively the goal of these trees is to maintain only the strong correlations present in the matrix. These MSTs tend to have a scale free [12] structure, with hubs (nodes with a large number of edges) occurring far more than would be expected from a random graph [267] [30] and a low average path length.

2.2.2 MST Forest

The MST is only unique if all edges in a graph have unique values. If not, there can be multiple MSTs present in the network. Djauhari proposes a method to detect if there are multiple MSTs in a network [69] and in a later paper [70], proposes a method by which the optimal MST from this forest can be selected.

To calculate this forest we iteratively modify the distance matrix using

$$D_{ij}^{k} = \min m(D_i, D_j^{k-1}) \tag{2.4}$$

where $m(x, y)$ is the element-wise maximum of x and y, D_i is the ith column of the distance matrix and min gives the minimum of the resulting vector.

Taking the original distance matrix D their algorithm proceeds as follows

1. Set $k = 0$
2. $D_0 = D$

3. $k = k + 1$
4. Compute D^k using (2.4)
5. If $D^k \neq D^{k-1}$ go back to 3
6. Calculate the matrix Δ as follows

$$\Delta_{ij} = \begin{cases} 1 & \text{if } D^k_{ij} - D_{ij} = 0 \text{ and } i \neq j \\ 0 & \text{if } D^k_{ij} - D_{ij} \neq 0 \text{ or } i = j \end{cases} \tag{2.5}$$

This Δ matrix then is the adjacency matrix of the MST Forest, and can be used along with the original correlation matrix to construct a weighted MST.

In practice, due to difficulties with floating point equality, correlation matrices inferred from stock returns tend to have unique edges. To overcome this, the edge weights must be rounded to a set number of decimal places if the MST forest is desired.

2.2.3 Average Linkage Minimum Spanning Tree

Constructing a minimum spanning tree using Kruskal's algorithm is very similar to a single linkage clustering method. In fact the algorithm to construct a minimum spanning tree can be written in such a way to look like a single linkage clustering algorithm. Tumminello et al. [263] show this, and furthermore how average linkage MSTs (ALMSTs) can be constructed. The algorithm to construct an ALMST proceeds as follows

1. Initialise the tree as a disconnected graph made up of the nodes in the input correlation matrix C
2. Create a new matrix $Q = C$
3. Find the edge with the highest weight in Q between different components S_h and S_k of the MST
4. Find the edge C_{hk} with the highest weight in the original correlation matrix C between S_h and S_k
5. Add this edge C_{hk} to the MST
6. Redefine Q as follows

$$Q_{ij} = \begin{cases} Q_{hk} & \text{if } i \in S_h \text{ and } j \in S_k \\ \text{mean}(q_{pt}, p \in S \text{ and } t \in S_j, \text{ with } S_j \neq S) & \text{if } i \in S \text{ and } j \in S_j \\ Q_{ij} & \text{otherwise} \end{cases} \tag{2.6}$$

7. If the MST is still disconnected, goto 3

2.2.4 Planar Maximally Filtered Graph

The MST model discards many correlations, keeping only $p - 1$ out of a possible $p(p-1)/2$. An alternative, less restrictive model, is the Planar Maximally Filtered Graph (PMFG), developed by Tumminello et al. [260]. This model keeps $3(p-2)$ edges over the $p - 1$ of the MST. Construction is similar to that of the MST:

1. Initialize an empty graph G with the same nodes as the correlation matrix
2. Sort edges of the correlation matrix in descending order in a list
3. For each item in the sorted list
   (a) If the edge does not increase the genus of G, add it to the graph

A planar graph is a graph with a genus of 0 - it can be drawn on a 2d plane without any of the edges crossing. All the edges present in the MST will be present in the PMFG. Furthermore the PMFG forces that all nodes are either in a 3-clique or a 4-clique (a clique is a set of nodes that are all connected to each other - their subgraph is complete). The PMFG requires slightly more data than the MST to consistently create the same graph, but this is to be expected as it retains more information.

2.2.5 Triangulated Maximally Filtered Graph

A disadvantage of the PMFG is that the planarity check at each iteration is computationally expensive. For large correlation matrices it may not be possible. There has been another algorithm proposed to construct a maximal planar graph which is more computationally efficient than the PMFG, the Triangulated Maximally Filtered Graph (TMFG) [163]. Rather than checking whether the planarity constraint is violated when adding an edge, this method adds a node into the graph in such a way that the graph is guaranteed to be planar. This is achieved by adding the node as a clique into a triangular face. The algorithm to construct a TMFG from a correlation matrix proceeds as follows:

1. Find the four most correlated nodes in the network
2. Create a clique out of these, and create a list that holds the set of triangles currently present in the graph
3. While there are less than $3(p-2)$ edges in the graph
   (a) Find the node most correlated with a triangle in the graph
   (b) Add this node to the graph, connecting it with every node in the triangle
   (c) Update the list of triangles in the graph

2.2.6 Thresholding

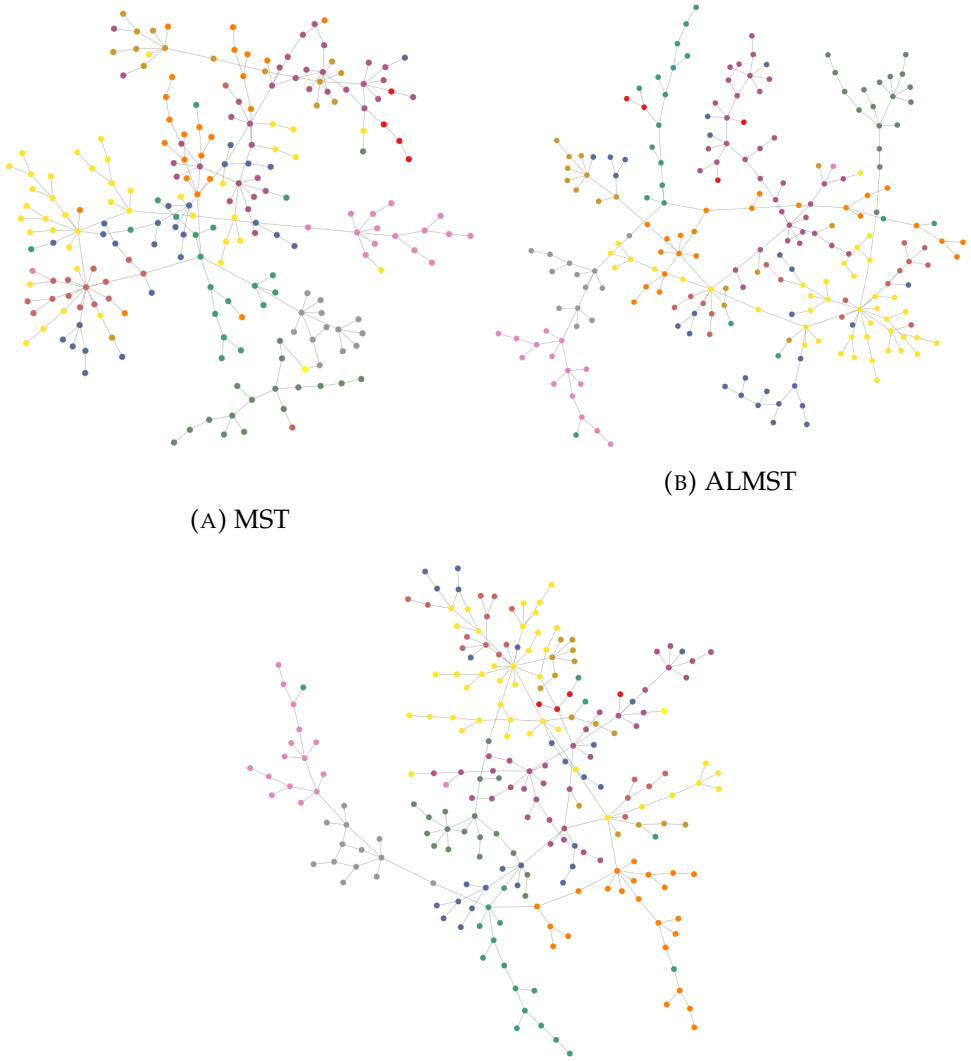
The topological methods have many attractive properties, but all throw away potentially large correlations that do not fit into their model. An alternative approach is to set some kind of threshold, where correlations above this threshold are set to 1 and those below set to 0. Often the absolute value of the correlation is chosen, to ensure large negative correlations are not discarded. In their seminal paper Onnela et al. [194] rank the correlation edges in terms of strength and add them one at a time to the graph, a model they name the “asset graph".

While thresholding does result in a robust graph [121], deciding on a threshold is non-trivial. Some authors chose an arbitrary value, but others calculate the mean ($\bar{C}$) and standard deviation (σ) of the correlation coefficents and set the threshold to be $\bar{C} + k\sigma$ where k is set to 1, 2 and 3 [191] [284]. Alternatively we can rank the edges in ascending order of p-value (for instance calculated by a statistical test, or by some form of bootstrap [105]) and set a threshold once the p-value increases over a set manner (usually 0.05).

2.2.7 k-Nearest Neighbours Graph

Another disadvantage of the thresholding model is that it can leave nodes isolated, which can be challenging to analyse. An alternative approach, termed the "k-Nearest Neighbours network" (kNN network) is to keep the same number of edges for each node [187]. This does require a decision for the value of k.

2.2.8 Examples



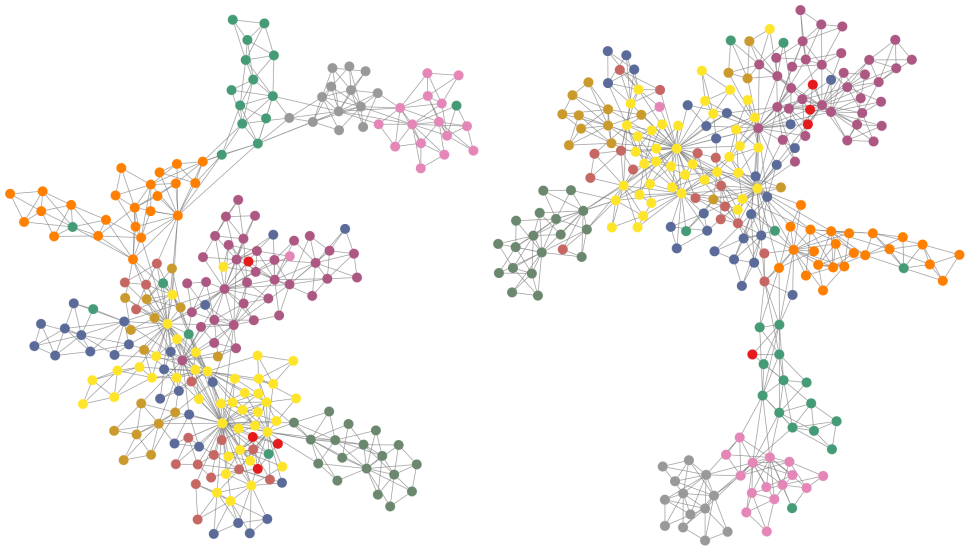
(A) MST

(B) ALMST

(C) MST Forest

FIGURE 2.1: Examples of the different types of MST generated from US stock returns. Nodes are coloured according to sector structure.

To provide a visual comparison of these methods, we construct some example graphs. These are generated from a correlation matrix inferred from 504 days of daily log returns data from 228 companies from the US S&P500 from 2000-01-03 to 2002-03-05. The networks are shown in Figures 2.1 (MST methods), 2.2 (planar methods) and 2.3 (kNN and Threshold). Nodes are coloured according to sector membership. The software used for this is described at the end of the paper.



(A) PMFG (B) TMFG

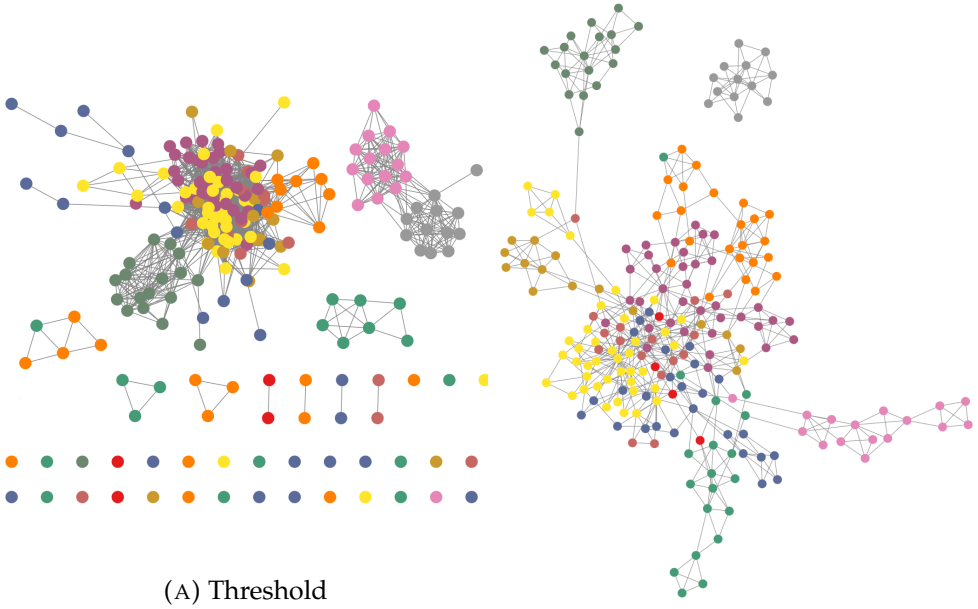
FIGURE 2.2: Example Planar Graphs generated from US stock returns

In all of the networks there is a significant amount of sector clustering, with nodes being very likely to share edges with those in the same sector. We can also see how the branches of the MST are likely to consist of nodes from a single sector. The three MST method are very similar, but there are differences between the ALMST and the MST, and cycles can be seen in the MST Forest.

In some ways the PMFG can be thought of a more permissive version of the MST (in does in fact always contain the MST [260] as a subset of the edge set) and this can be seen in this example - there are branches that contain companies in one sector in a similar manner to the MST, but these have more edges that can connect them to other parts of the graph. Similarly to the MST methods, the TMFG and PMFG do have quite similar structures.

The thresholded graph has split into multiple components, but usually the largest component is significantly larger than the others, and is densely connected. There is also a number of isolated nodes (nodes with no edges). In the kNN network there are no isolated nodes, but the network is not connected, with a number of different components being present.

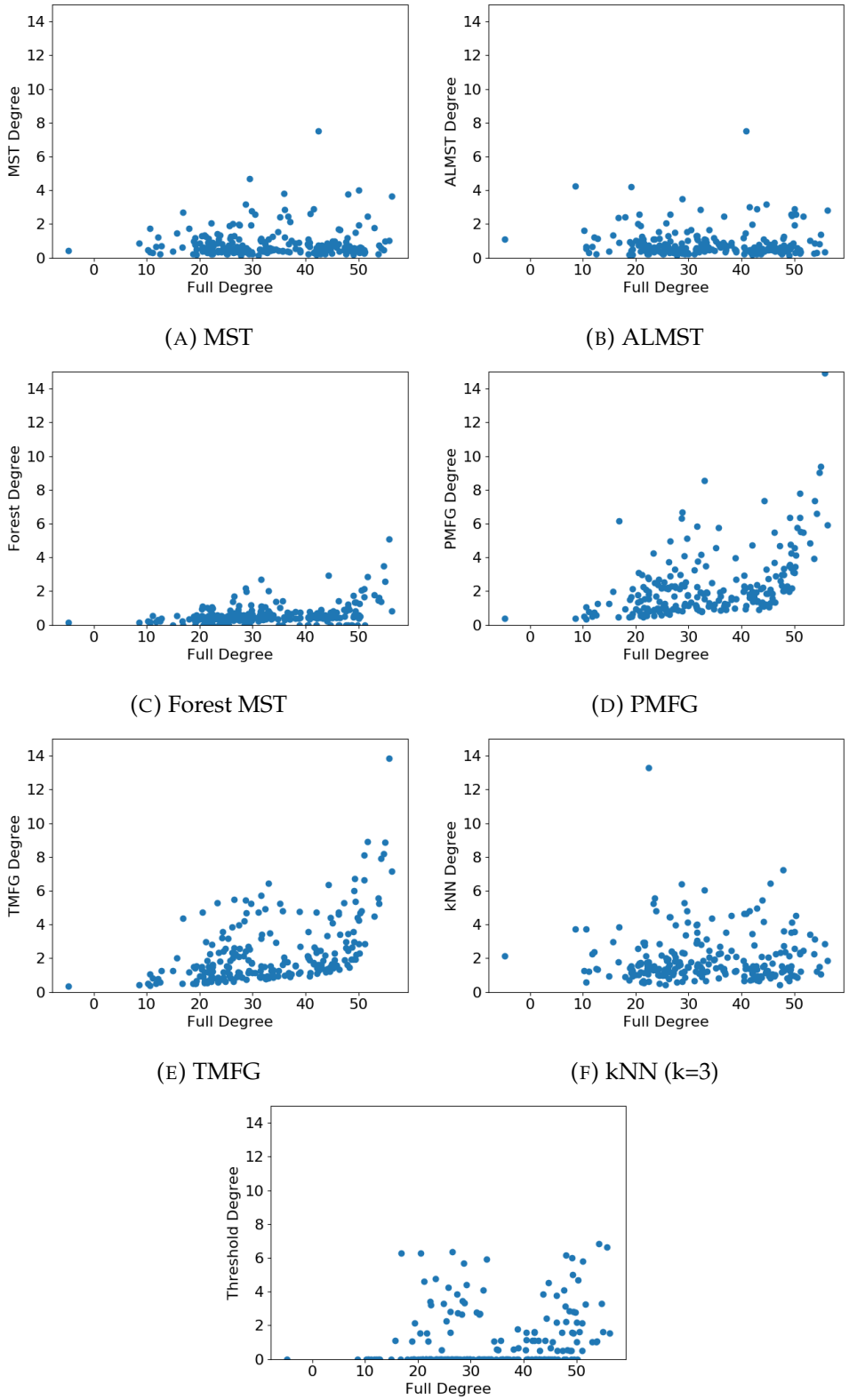
We then plot the weighted degree centrality of each node (i.e. the sum of the row that corresponds to the node in the adjacency matrix) in these filtered graphs vs that of the same node in the full correlation matrix in Figure 2.4. The filtration methods do tend to give a very different picture of the correlation network, with nodes being able to have a significantly different centrality in the filtered network compared to the full one. The methods that retain less edges (i.e. the MSTs) tend to have a less of a relationship between the degree centrality in the full vs filtered correlation matrices.



(A) Threshold

(B) kNN

FIGURE 2.3: Example of kNN and threshold networks generated from US stock returns



(A) MST

(B) ALMST

(C) Forest MST

(D) PMFG

(E) TMFG

(F) kNN (k=3)

(G) Threshold

FIGURE 2.4: Plot of the weighted degree centrality of a node in the full correlation graph (x-axis) to its weighted degree centrality in a filtered graph (y-axis) constructed using the example methods on US stock log returns. This shows the differences in the severity of the filtration methods, and how it they capture different information to that of the full graph.

2.2.9 Windowing

Financial returns are non-stationary, and this has an effect on networks inferred from them. Correlations tend to rise during periods of market stress and drop during periods of calm [211]. To solve this, we take a small window of data where we can assume the data is approximately stationary. Deciding on the size of this window is challenging, as there is no objective way of setting the size. It is usually arbitrarily set to either a round number (say 1000 days) or to a relevant size - for instance for daily returns one trading year is a common choice (there are 252 trading days in a trading year). There is also a trade off - larger windows mean there is less noise in the estimation of the correlation matrix, but a greater chance of non-stationarities affecting the results. Different window sizes also have significant affects on the structure of the networks [291] [238].

Potential solutions to this include change point detection, or using a form of weighting observations with an exponential decay, proposed by Pozzi et al. [209]. This can be thought of as like a soft threshold, compared to the hard threshold of a sliding window. By giving greater weight to observations close to our period of interest, and reducing the weight of those further from that period, we can gain a snapshot of the network at that moment in time, while avoiding potential discontinuities from hard thresholds. This still does require a decision on the decay parameter.

2.3 Correlation Network Analysis of Financial Markets

In this section we discuss how these filtration methods have been used to analyze financial markets. We focus specifically on stock returns, and discuss applications to other assets later in Section 2.6.

2.3.1 Minimum Spanning Tree Analysis

In [196] Onnela et al. study MSTs inferred over time using a sliding window. Their data covers the 'Black Monday' (October 19, 1987) crash, allowing us to understand how these trees behave during times of stress. Firstly they find that the MSTs tend to be relatively well preserved over time, with around 80% of the edges being maintained from one MST to the next. For the edges that do disappear, they tend to disappear relatively quickly (i.e. within a small number of time windows), although there are some edges that are maintained for over a decade. The most central node is taken by a relatively small number of companies, with their dominance maintained for most of the dataset. The trees become less stable during times of market stress, with large changes in the structure occurring. Furthermore the tree tends to shrink in diameter as the companies start to behave more similarly due to the macroeconomic effects of the crash. Finally they look at the consequences for portfolio optimization, specifically with Markowitz portfolios. Generally companies that have large weights in these portfolios tend to be on the leaves of the trees. There is however evidence that this is merely an empirical observation, with algebraic analysis showing no reason that this should be the case [111].

Wilinski et al. [283] further show that during times of market crisis there are significant structural changes in MSTs, transitioning from a scale free MST with a hierarchy of local hubs to being orientated around a single super hub.

MST based models have been applied to other markets, including the Japanese [120], UK [56], Chinese [293], South African [159], Turkish [122], Italian [58], Korean [119], Polish [89], Indian [236], German [37], Malaysian [7] and Brazilian [247].

Generally the US has a massive influence on these markets which is not reciprocated [78], due partially to its economy being significantly larger than any other. There is also evidence to suggest emerging market are less efficient and are structured differently to mature, more liquid markets [192] [164] [285]. However the shortening of the MSTs during times of market stress is seen in other countries as well as the US [145].

Nie et al. [189] analyse the structure of MSTs using Renyi indices, a measure of the heterogeneity of a system. A smaller Renyi index implies a graph has nodes of a similar degree, while a larger Renyi index implies a graph with nodes that have larger differences in degree. In general they find that the Renyi index of MSTs constructed from the S&P500 tends to move in the opposite direction to the index. This implies that the MSTs have more nodes of a similar degree during times when the market is increasing in value, and that there is more variance in the node degrees when the market is crashing. On an international level, the Renyi index of the Chinese market is more volatile than that of the US markets.

In [190] Nobi et al. use a measure of the hierarchy present in a network (a measure defined by [259]), defined as the fraction of paths between pairs of nodes that are degree hierarchical. For a path from node i to node j to be degree hierarchical, node i must have a smaller degree than node j. They find that the hierarchy in a financial MST constructed from S&P500 returns tends to rise just before a crash, and falls during the financial crisis itself.

MSTs have also been applied in the neuroscience field [254]. Notable examples include Vertes et al. [270], who compare neuroscience networks and those inferred from stock returns. Applying both an MST model and the asset graph to correlations inferred from fMRI data and from stock returns from the NYSE (although they notably do not have any crisis periods in their dataset), they find both networks show small world characteristics, with a high level of clustering, high degrees of modularity and fat tailed degree distributions. Brain networks are shown to be slightly less clustered, but more robust than their financial counterparts. They have also been used to compare the differences in brain structure between controls and patients with Multiple Sclerosis [253].

2.3.2 Thresholding Analysis

In their seminal paper, Onnela et al. [195] analyze the asset graph. Firstly they show these graphs are different from a MST, only sharing around 25% of the same edges when they have the same number of edges. Comparing this to an Erdos-Renyi random graph with an equal chance of an edge existing between two nodes, they show these asset graphs have properties that are significantly different from said random graphs. Most notably the asset graphs are significantly more clustered than a random graph, for instance at a connection probability of 0.2 they have a mean clustering coefficient of around 0.7 compared to 0.2 for that of the random graph. The behaviour of the clusters in the asset graph is very different compared to the random one (for instance there is a maximum of 9 clusters in the asset graph compared to 72 for the random one). Their final note is that after a connection probability of around 0.3 most of the edges added seem to have an effect similar to that of a random graph, indicating they may be noise.

Boginski et al. [26] further study these asset graphs. Experimenting with setting various thresholds, they study how this then effects the degree distribution of the graphs created. They claim that with a sufficiently high threshold the degree distribution follows a power law. With these asset graphs they then formulate maximum cliques

and maximum independent sets. A maximum clique is a set of nodes that are all connected and cannot have another node from the graph added to the clique without the set no longer being a clique. A maximum independent set is the opposite - a set of nodes from the graph that are not connected and cannot have another node added to them without breaking this property. The interpretation of these maximum independent sets from a financial perspective is that theoretically they are a set of diversified stocks. They can be found relatively efficiently and in a future paper [27] (reviewed below in Section 2.7) the authors evaluate portfolios constructed using them.

As well as to the US markets, other authors have applied this model to the Russian [268], Chinese [108] [232], British [55], Iranian [76], German [24] and Greek [68] markets.

Souto et al. [243] look at relating the VIX index to the average degree of thresholded correlation matrices, and to the presence of balanced or unbalanced triad motifs in a signed correlation graph (setting edges below a threshold to -1 and those above to +1). They find significant correlation between the VIX index and the average degree and that the presence of balanced positive triad motifs is a good predictor of the VIX index.

Heiberger [102] studies correlation networks with a set threshold during times of stress and calm. Using the May-Wigner theory of network stability from ecology, they claim the markets become less stable during times of market stress, and that the modularity tends to drop too.

An interesting piece of analysis is performed by Xia et al. [284], who compare thresholded networks inferred from Chinese stock returns. In particular they compare how the structure of the market changes from the financial crisis, to a period of growth, and then to a local stock market crash. The threshold is set using the mean correlation plus a multiple of the standard deviation. While the GDP of China grew during the financial crisis, it still affected the country [149], and these effects are visible in the stock market network. Both the mean correlation and mean degree of the thresholded network are higher than during a time of calm. However the local stock market crash has a far more drastic effect, with the mean correlation rising significantly, as with the mean degree, even though the threshold is higher. The modularity of the network also rises during these crisis times. It is interesting to see how much more severe the effect of this local crash is against the global one.

Similar work has been done studying the effects of the financial crisis on the Korean stock market [191] and the effects of Brexit on the British markets [8].

2.3.3 PMFG Analysis

In the same paper where Tumminello et al. [260] propose the PMFG, they also demonstrate its properties on returns from the NYSE and compare them to those of an MST. They find that the cliques tend to be composed mostly of companies belonging to the same economic sector. Evaluating the stability of the PMFG vs the MST by varying the sample size of the data, they show that the PMFG requires more data than the MST to be stable.

In a similar manner to Onnela et al. [196], Aste et al. [5] study how the topology of PMFGs inferred from stock returns varies over time. Firstly they note that 96% of the correlation coefficients retained by the PMFG are significant, compared to 100% of those in the MST. This seems reasonable, as the PMFG maintains more edges than the MST, so we would expect a slightly higher number of noisy edges to be found. To study the topological change in a graph they use a measure known as the

'T1 Distance', which quantifies the minimum distance between two nodes that were neighbours in the first graph, but may become separated by longer paths in future graphs. By this measure, the PMFGs can be considered more stable than the MSTs. However if the size of the networks is taken into account, the two seem to have a similar level of relative stability. Finally they explore how the centrality of the financial sector has changed using betweenness centrality, eccentricity, degree and closeness. In general the financial sector is quite central in the US markets, which has been found in many other studies. During the financial crisis of 2008/2009, the centrality of the financial sector increases, but there is evidence that overall it is decreasing over the 13 year period (1996-2009) that they consider.

Di Matteo et al. [65] further explore the similarities and differences between PMFGs and MSTs. Constructing PMFGs and MSTs from the same dataset of NYSE daily returns, they then calculate centrality and periphery measures for each node in the filtered networks. Comparing these measures between the MSTs and the PMFGs, they find that the two networks tend to agree on sector centrality, with both finding the financial sector to have the most companies in the top 50 most central nodes. Using these measures to then cluster the nodes, they find that the nodes can be split into 6 clusters of varying connectivity and centrality.

Musmeci et al. [178] study the persistence of structure in correlation matrices from US and UK financial returns. They quantify this using two measures, meta-correlation (the correlation between the entries of adjacent correlation matrices) and edge persistence (the fraction of edges that are shared between adjacent PMFGs). Particularly interestingly they take multiple networks into account for these measures, and use exponential weighting to ensure that results from networks closer in time are given a greater weight than those further away. Furthermore they also develop a measure q which is the ratio of the amount of volatility present in the current window vs the next one, and look at predicting this from the edge persistence and the meta-correlation. They show that this can be predicted, with an accuracy usually above 50%, and in some cases up to 83%. This shows how financial networks may be useful for predicting future volatility, which could be useful when deciding how or when to rebalance a portfolio to reduce risk.

Zhao et al. [291] perform a detailed study of PMFG structure during two financial crises, the dot-com bubble (1999 - 2002) and the subprime crisis (2007 - 2009) using returns from the S&P500. Firstly they investigate the consequences of choosing different window sizes for the structure of the networks, and show this choice has a significant effect on the structure of the network. Next they take a novel approach to understanding the changes in structure in the markets by using the entire dataset up to the particular day in question, rather than using a window of constant size. With this method they still find that the structure of the network changes during times of crisis, with the average shortest path length decreasing, the average clustering coefficient increasing and the heterogeneity of the network (as defined by [77]) decreasing (indicating the companies become more similar). They also find that these PMFGs constructed from the full dataset show a greater degree of sector clustering compared to those constructed using windows, even when the windows are large. Using InfoMap [220] to perform community detection, they show the modularity increases during these crisis periods, and that the number of intersector edges drops during both crises and stays low once it has dropped, indicating a permanent increase in sector correlation. Finally they study the edit distance between adjacent PMFGs in time. This spikes before a crisis, indicating a large change in structure.

Nie et al. [188] explore the Renyi index and dimension of MSTs and PMFGs constructed from the US, UK and Chinese markets. The dimension of a network is

the exponent between the volume and distance in a network [228]. The volume of a node is the number of nodes that are within a given distance of it. The volume of a network is the mean of the node volume for a given distance. To calculate the dimension, the volume of the network is calculated for a set of distances, and the value of the exponent describing the power law relationship between distance and volume is estimated. The authors show there is strong negative correlation between the Renyi index and dimension (-0.95 Pearson correlation) for both the MSTs and the PMFGs for all countries. This shows that as the network shrinks in diameter (which is common during a financial crash as correlations increase), it becomes less heterogeneous, and more oriented around a few nodes.

Yan et al. [286] study the efficiency and robustness of MSTs and PMFGs. In particular they focus on the effects that node removal has on the structure of the network. This is measured using the change in efficiency and by the fraction of nodes that are placed into a separate component when a node is removed. Efficiency is calculated as the sum of the inverse of the shortest paths between nodes. Two methods of node removal are investigated, firstly removing the most connected nodes and secondly by randomly removing nodes. Perhaps unsurprisingly, removing the most connected nodes has a drastic effect on the MST, causing a large drop in the efficiency and causes a large fraction of the nodes to be placed into a different component. A similar effect is seen for the PMFG, which is more unexpected. However if random nodes are removed, the drop in efficiency is much lower, indicating that it is only a few nodes that play important roles in PMFGs and MSTs, similarly to other small world networks [3].

PMFGs have also seem applications in fields other than finance, notably with gene expression data [242].

2.3.4 kNN Network

The kNN network has only been studied in the original paper where it was proposed. In this paper, Nie et al. [187] find that nodes that have a large entry in the leading eigenvector of the full correlation matrix tend to have a high degree centrality in the kNN network. Secondly they use Newman's spectral algorithm to perform community detection on the kNN network. Varying the values of k, using modularity as a measure they suggest $k = 3$ as an acceptable value that provides good community separation while giving a sensible number of communities (in this case 11). Studying the correlation matrix of each community, they find it has a eigenvalue structure that is different to what would be expected using RMT. kNN networks are however not guaranteed to be connected, and in fact can require reasonably large values of k for this to occur. kNN networks have also been applied to gene expression data [52] for the study of cancer.

2.3.5 Varying Time Period

Mostly the focus for the inference of these networks is on daily data. Correlations tend to increase as we increase the time period (known as the Epps effect [74]), and data tends to become less noisy. Institutional investors also tend to hold portfolios for long periods of time, so these longer term correlations are important to understand. However larger time windows mean less data is available. Furthermore we may be interested in the dynamics of intraday data, with high frequency trading becoming very common. Therefore studying how these correlations and networks vary for different time periods is important.

On this theme, Tumminello et al. [261] look at how PMFGs are affected by these different time periods. In particular they study how the sector structure evolves as the time period increases by looking at the average number of intra-sector edges per sector (i.e. how many edges a company in a sector has to companies in the same sector) The sector structure emerges a quite small time horizons for some sectors - even at 5 minute intervals the financial, utility and energy sectors have many intra-sector edges. For other sectors this takes longer periods to emerge, for instance, the consumer cyclical and healthcare sectors.

A different approach is taken by Wang et al. [276], who propose to use wavelet methods to decompose daily returns into various frequencies. This allows us to get views of correlations over longer time periods while still making use of the availability of daily data. Once these multiscale correlation matrices are created, they construct MSTs and PMFGs and study their various properties. In these matrices, the mean correlation tends to decrease as time periods increase, but the standard deviation increases with the increased time period, as there are still some large correlations present. For the shorter time scales, sector clustering tends to occur in the MSTs, but at the largest scale (which is on the time scale of 256-512 days) the sector structure almost completely disappears. Looking at the degree centrality in these trees, companies from the Financials, Materials and Industrials tend to make up at least one of the top 3 most central nodes in all of the time scales, showing their importance in the US stock markets. Finally the authors perform community detection on the PMFGs constructed from the dataset. In a similar fashion to the MSTs, the sector structure becomes less pronounced as the time scale increases, as both the number of clusters increases and the sector make up of each cluster becomes more diverse.

2.3.6 Network Structure

While many authors have claimed their inferred correlation matrices have a small world character, evidence from the neuroscience world has shown this may be due to a bias in these correlation matrices rather then necessarily true clustering [289] (although since there is not a true model underlying the system this is difficult to actually define). Effectively the argument is that if there is a direct correlation between nodes a and b, and a is correlated with c, b is also likely to share some correlation with c due to the definition of the measure. The relationship between b and c is defined as an indirect correlation. This therefore increases the clustering coefficient in networks constructed from these matrices purely due to these indirect correlations rather than due to true clustering. The authors propose to solve this by using more appropriate null models for correlation based networks rather than using a random graph. One of their examples is to use a randomly generated correlation matrix with the same distribution of entries as the observed one. Using this, they show that this correlation based null model has a clustering coefficient of 4 times that of a random graph. Interestingly though this does not apply for partial correlation models, in fact they are less clustered on average than a random graph, although in some ways this could be considered the goal of partial correlation based models - to remove these indirect correlations. This has also been noticed by Hlinka et al. [104] in the context of financial networks.

2.3.7 Lead-Lag Networks

The efficient market hypothesis is a hypothesis in finance which states that stock prices reflect all known information. It implies that future returns are independent from past returns. Simple statistical tests reveal this to not be the case [154]. This therefore implies there should be some cross-correlation between stock returns. It is generally accepted there is little autocorrelation between stock returns [251], but there are significant cross-correlations between stocks. A lead-lag network is constructed using cross-correlation between asset returns.

Kullmann et al. [137] create directed networks from lagged correlations using 100 second return data. They find that more frequently traded companies tend to have a lot of influence, and that smaller companies tend to influenced by many larger companies while a larger companies influences many small ones. Other authors have applied the same model to the Chinese market [91].

Curme et al. [61] take a more detailed study. They infer lagged correlations between stock returns from companies listed on the NYSE, sampled at 5, 15, 30, 65 and 130 minutes and analyze the results. These correlations are then checked for significance using a bootstrap method, and any insignificant correlations are discarded. In general it appears that correlations increase as the time period decreases - the momentum of price movement is clearly important, and that lagged correlations disappear for timescales longer than an hour. Most of the correlations are positive rather than negative. Furthermore the number of significant correlations has dropped over time, implying that markets have become more efficient. Since these are lagged measures, a directed network can be inferred from the results. In these networks most of the nodes have a small out degree and a large in degree - they influence few other nodes but are influenced by many other nodes.

2.3.8 Bootstrapping and MST Reliability

If we wish to use these MSTs in applications, we must study their consistency and reliability. Tumminello et al. [263] use the bootstrap [35] to study how well edges are maintained in MSTs. They create minimum spanning trees using returns from 300 stocks from the NYSE and use a bootstrap to see how reliable the edges in these minimum spanning trees are. By selecting only specific returns from the dataset they check if the same edges are present in the MST constructed from this smaller dataset compared to the full one. Furthermore they compare the robustness of the ALMST to the normal MST. In general they find the ALMST has a slightly higher success rate in linking companies in the same sector although 85% of the edges between the two MSTs are shared. For the reliability they find that the average bootstrap value for the MST is 0.627 and 0.602 for the ALMST (i.e. on average, the edges in the MST inferred from the entire dataset are present in 62.7% of the bootstrapped MSTs). Edges between companies in the same sector tend to be present slightly more often (0.740 for the MST and 0.725 for the ALMST) and edges between companies in different sectors tend to be present less often (0.469 for the MST and 0.426 for the ALMST). Furthermore the strength of an edge does not necessarily mean it will be selected often - there is little relationship between the strength of an edge and its bootstrap value. In a further paper [262], the Kullback-Leibler distance and bootstrapping are used to compare the stability and amount of information retained by various correlation filtration methods. They find that single linkage clustering is the most stable, but removes the most information, while the average linkage clustering retains more information at the cost of stability. Taking a random matrix

theory approach (which is described later in section 2.5) retains the most information at the cost of the lowest stability.

An alternative bootstrapping procedure is performed by Musciotto et al. [177], who again use this to analyze the robustness of MSTs. In their work they test two different bootstraps, one using pairwise comparisons (bootstrap on each pair to calculate a distribution for the correlation between two assets) to one on the overall model (referred to as the row-wise bootstrap). They find that both show the MSTs are relatively robust, generally selecting similar edges on the various bootstrapped datasets. However the pairwise bootstrap is much more restrictive than the row-wise one, with edges usually having a lower p-value (i.e. appearing in less of the MSTs inferred from the bootstrapped dataset) and is not guaranteed to produce a positive-definite correlation matrix. To quantify this, if we select a p-value of 0.2 almost all the edges in the MST on the total dataset are selected for the row wise bootstrap, but for the pairwise bootstrap this drops to 85%, and this disparity only increases as we increase the threshold for the p-value. It does however gain a better estimate of the largest eigenvalue and the MST produced has a higher success in discovering intra-sector relationships, so the authors argue it is a tool that could be useful to gain a slightly different perspective when constructing MSTs. They suggest that edges that appear in over 20% of the bootstrapped MSTs should be considered informative and reliable.

Kalyagin et al. [121] study the reliability of MSTs, PMFGs and thresholding in a similar manner. Firstly they estimate the correlation matrix of a set of US financial returns, and construct a filtered network from this (i.e. a MST, PMFG or thresholded graph). This filtered network is then treated as the reference graph. Next samples are drawn from a multivariate normal distribution with the same correlation matrix. Filtered networks are then again constructed from these new correlation matrices and then compared to the reference model. This is used to quantify the uncertainty present for each method for a given number of samples. Both the MST and PMFG required a large number of samples (i.e. over 10,000) to achieve a low uncertainty (i.e. a low edge difference from the reference network) The thresholded networks are more robust, requiring only 300 samples to achieve a similar level of uncertainty.

2.4 Alternatives to Pearson Correlation

Pearson correlation tends to be the most ubiquitous method used to infer the relationships between assets. However it assumes the data is normally distributed and only measures linear relationships, neither of which can be desirable when analyzing financial data. There are however alternative methods that can give us a different view on how two assets may be related. While there may be many goals in using alternative methods, we divide them into broad categories - those which remove the effects of other stocks on relationships, those which capture non linear relationships and those which filter the returns to remove noise.

2.4.1 Methods to Remove Other Relationships

Firstly we focus on the methods that remove other relationships. Two stocks can be correlated due to sharing a common factor driving them rather than because they are in fact influencing each other. Perhaps the simplest method to attempt to remove other relationships or market modes is to subtract the mean return across the entire market from each individual stock return [131]. The network is then inferred by

calculating correlations between these new return time series. This however only removes overall effects.

Partial correlation can be a way of solving this issue, with the advantage that it can remove common factors between groups of stocks as well as the market as a whole. A univariate partial correlation model is applied by Kenett et al. [126] on US stocks. This method is generally referred to as a 'dependency network'. By removing the effects of one stock on the correlation between another two, they create directed networks from stock returns in an attempt to understand influences in the US economy. Their main conclusion is that the financial sector has a disproportionate effect on the US economy, as it has a large out degree in these networks. They further extend this model in two more studies [125] [128]. In [125] they construct MSTs and PMFGs from these dependency networks from windows of return data and compare how the topology of the networks change using a measure of divergence and the average clustering coefficient and path length. They find that the dependency networks have peaks in the average shortest path and average clustering coefficient during times of crisis, and that during these crisis there is firstly a large change and the networks are not very similar, followed by the networks becoming significantly more similar during these times. In [128] they perform statistical tests to remove the non-significant connections. Quantifying the stability of various countries using returns from their index by ranking the stocks on out degree and using the Kendall's τ coefficient to measure how preserved this is through time, they show that the more developed nations, such as the US, UK and Japan have relatively stable stock markets, but this is not the case for India. Looking at sector specific effects, they find the financial and energy sectors tend to connected within themselves i.e. they mostly influence themselves while the Industrial and Materials sectors are much more influenced by other sectors.

A more complete form of partial correlation, calculated by inverting and scaling the correlation matrix is applied to index returns by Wang et al. [277]. With these networks they construct MSTs and compare and contrast the edges. They find that both trees have power law degree distributions, and that the correlation based tree is much more tightly clustered than the partial correlation based tree. Furthermore the partial correlation based tree reveals that the US, Germany and Japan serve as hubs in the world stock markets, structure which is not present in the correlation based networks. Similar models have been applied to the US [174] and to the Mexican [258] markets.

Interestingly there have been methods developed to efficiently construct partial correlation MSTs and TMFGs. Barfuss et al. [13] propose a method they name 'LoGo' to efficiently construct partial correlation networks from MSTs and TMFGs without having to invert the full matrix. They achieve this by firstly constructing the filtered correlation matrix with the preferred method (i.e. MST or TMFG) and then by using local inversions to calculate the partial correlation for each of the edges.

An alternative method of normalizing the correlation matrix to reduce the effect of the overall market is demonstrated by Kenett et al. in [127]. They use a method proposed by [15] in which they create a meta-correlation by measuring the Pearson correlation between rows and then multiplying this by the raw correlation value. This new matrix is termed the affinity matrix. By removing the background market effect, this allows a greater ease in detecting clusters of stocks, an effect noted in other research [181]. Furthermore they measure the amount of information in these networks by looking at the dispersion of the eigenvalues compared to the standard correlation networks. The standard correlation networks have a strong time varying measure, where periods of market distress are clearly visible. In this new measure,

the is no real time varying measure of information available, showing the overall market effect is removed. In further work the spectrum of these affinity matrices is compared to that of partial correlation and normal correlation to evaluate the amount of information available. In [124] Kenett et al. study the eigenvalues and eigenvectors of 3 different types of correlation matrix, a raw one, their affinity matrix and a partial correlation based method, where the stock index is removed from the market. Their main finding is that this partial correlation matrix has a much larger number of information bearing eigenvalues and eigenvectors than the other two correlation methods. Furthermore the affinity matrix tends to have a much larger largest eigenvalue than the other two methods and tends to contain the largest amount of information measured using the spectral entropy. This model is used in a manner by Shapira et al. [230] to study how the index influences the rest of the market. By removing its effect using partial correlation and the affinity matrix methods ([15]) they show that the index is responsible for most of the correlations between stocks.

2.4.2 Quantifying non-linear relationships

Rank correlation methods can be a simple way of quantifying non-linear relationships between assets. They also have the advantage of being robust to outliers, as we compare the ranks of the variables rather than their values. Spearman correlation based financial networks have been studied by Shirokikh et al. [234] although they did not compare their networks to Pearson based ones, instead looking at extracting k-cores. A more thorough comparison is performed by Pozzi et al. [209] who compare Kendall's τ and Pearson correlation in terms of information content using random matrix theory. Generally they find that Kendall's τ tends to hold information better than Pearson correlation, with more non-noise eigenvalues present. The authors also propose a method to solve the issues of choosing a window size, which is mentioned in Section 2.2.1

Another study into the differences between various correlation methods is performed by Musmeci et al. [180], who take a multilayer network approach and infer the layers using Pearson correlation, Kendall's τ correlation, partial correlation in the manner of Kenett ([125]) and tail correlation. These are filtered using the PMFG method. In general the layers appear to have different structures, with between 30% - 70% of the edges being unique to a particular layer, but this is sector dependent. The Pearson and Kendall τ correlation tend to agree the most on the degree of nodes, with a node degree correlation of around 0.7, but this drops to 0.5 for the other measures. Despite this, the layers are all affected by macroeconomic effects, with market crashes tending to cause an increase in layer weight, but also a decrease in edge overlap (i.e. less agreement between the layers)

Mutual information is another potential solution in attempting to identify these non-linear relationships. Fiedor [80] uses mutual information and the mutual information rate to quantify similarity between stock returns. Constructing both MSTs and PMFGs from the mutual information between stocks they find these are significantly different from the ones constructed using correlation. To measure this they look at various network measures, notable node degree, where the correlation between the degree of nodes inferred from the correlation base MST and the mutual information based MST is 69%, and 67% for the PMFG, implying that around a 3rd of the relationships in their dataset (the NYSE100 index) contain non-linear dynamics.

This is further extended to lag lead relationships in a following paper by the author [80]. In this they infer directed networks using mutual information from

log returns at 1 minute intervals from the NYSE100, and remove any edges that are not statistically verified using a bootstrap method. They find that the number of significant edges decreases more rapidly as time periods increase with mutual information based networks than with correlation based networks, and that mutual information finds more edges within a sector. Finally they also discover lagged relationships with lagged daily returns, indicating that the efficient market hypothesis is not strong.

The differences between Pearson correlation and mutual information based networks are further studied by Hartman et al. [100]. They find that the correlation between edge weights of mutual information and Pearson correlation networks is 0.87, but if the individual stock return distributions are processed to ensure they are Gaussian, this correlation jumps to 0.9775, indicating that most of the differences in these networks is due to marginal non-Gaussianity rather than non-linearity. However there are still some significant non-linear effects that can occur, particularly during financial crises.

While again much of the focus on these mutual information networks has been on US data, it has also been applied to Indian [231], Polish [81], Chinese [95] [288] and Australian [287] stock markets.

Once inferred, these lag lead mutual information networks have been applied to attempt to predict the VIX index by Begusic et al. [17]. Using transfer entropy, which quantifies the amount of uncertainty in a variable X is reduced by including the past of Y, they create a directed network and look for the existence of strongly connected components to find cycles of feedback. They find that the number of relationships increases during times of market stress, indicating a more chaotic and complicated system, and that the predictability of US stocks decreases as time has gone on, possibly due to the increased liquidity of the markets. Applying their network to predicting the volatility of the S&P500 index 1 year in advance they find their 'Strongly Connected Component Index' increases the R^2 measure compared with purely using the VIX index from 0.29 to 0.45.

A disadvantage of information theoretic measures is that they require us to have a distribution of the data. Binning is a usual tool to solve this, but due to the high dimensionality of the data and the desirability to have a small window (due to non-stationarity of the data) we can only use a small number of bins to get an acceptable distribution.

2.4.3 Filtration Methods

We define filtration methods to be attempts to remove the effect of outliers or volatility while still using Pearson correlation on the results. These do not necessarily quantify non-linear relationships, but instead remove the effects that outliers or non-linearities may have on the data.

A simple but effective method of dealing with non-linearities is symbolisation. This involves converting a time series into a set of integers, which allows easier analysis of noisy complex systems. For financial returns this can usually be done by mapping a return to where it falls in a set of regions. For instance if we choose two symbols, a return could be mapped to 1 if it below the mean of the entire series, or 2 if it is above the mean. The distance between these two new series is then calculated, and then an MST, PMFG or other desired filtration method can be used. Furthermore another attractive property of this method is that it can be used to combine multiple pieces of information about an asset - say both the return and the volume traded

of a stock. Such methods have been applied to the Italian stock markets [41] [40], currencies [38] and to the Euro Stoxx constituents [39].

Isogai [112] uses a DCC-GARCH model for data filtration. If a large market shock occurs the increased volatility could distort our measurements of correlation within the markets. DCC-GARCH allows volatility to be time dependent, and we can use it to remove the effects of volatility on these stock returns. Using Japanese stock returns the authors show that these DCC-GARCH based correlation models detect periods of change well but successfully reduce the amount of correlation present once this shock has occurred.

Another method is Detrended Cross-Correlation Analysis. This allows the quantification of correlations for non-stationary data, which is clearly an attractive property when dealing with financial returns. Originally proposed by Podobnik and Stanley [207], it has seen many applications in currency analysis [275] [279] and in analysis of both the US [278] and Chinese stock markets [51].

Khoojine and Han [130] fit an autoregressive model on stock returns from the Chinese markets, and then use the results from the models to construct mutual information based TMFGs and MSTs. This model is used to study the Chinese stock market during 2015 and 2016, which covers a period of market disruption.

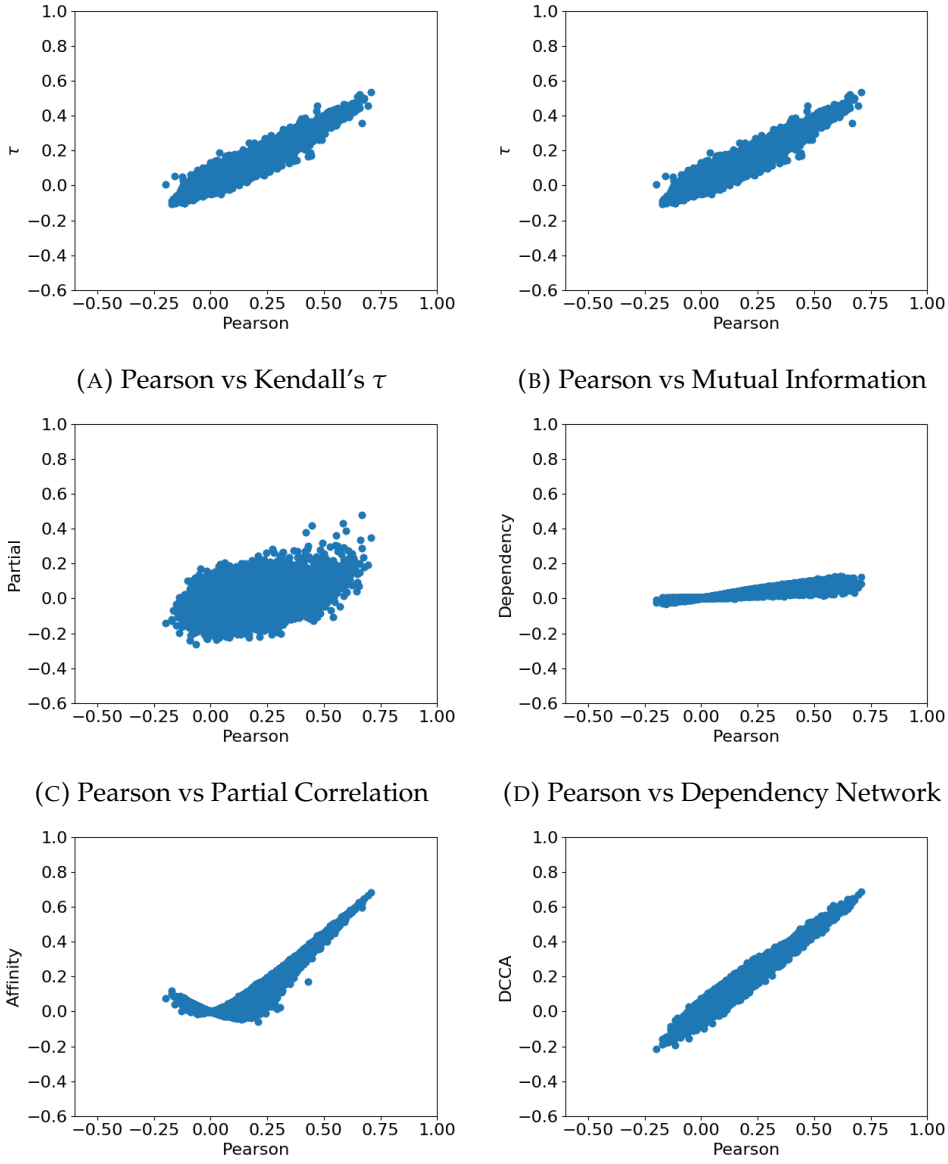
2.4.4 Other

There are a few methods we have discovered that do not fit into these categories. The first is based on Escoufier's RV coefficient [216], which is a multivariate extension of Pearson correlation. While close price is the most commonly used piece of information about a stock, other prices are available too - open, high and low. Lee et al. [144] use the RV coefficient to combine these pieces of information and construct MSTs to analyze Malaysian stock returns. In a further paper they extend this method to using the MST Forest on NYSE stock returns [90].

The second is the use of causal networks. Causality in itself is very challenging to measure, but we can approximate it using Granger causality. This is applied to the monthly returns of banks, insurance companies, brokers/dealers and hedge funds to attempt to quantify the amount of systematic risk in the system by Billio et al. [23]. A formal definition of systematic risk is *any set of circumstances that threaten the stability of or public confidence in the financial system*. This can be related back to interconnectedness - if one bank fails, are the rest likely to? (or are the public likely to believe the rest will, which could amount to the same thing) They find that in the build up to, and during, the financial crisis of 2008, there was a significant build up of relationships in these Granger causality networks, showing the increase of interconnectedness which ended up helping to exacerbate the crisis. Furthermore they find that banks tended to have far more out links that the other groups, indicating that banks play a key role in transmitting shocks to the rest of the financial sector.

2.4.5 Examples

Here, we show how these various methods compare. In a similar manner to Section 2.2.8, we use log returns from the US S&P500, and use each method to calculate similarity matrices between each asset. We then show scatter plots of how various alternatives compare to Pearson correlation in Figure 2.5.



(A) Pearson vs Kendall's τ

(B) Pearson vs Mutual Information

(C) Pearson vs Partial Correlation

(D) Pearson vs Dependency Network

(E) Pearson vs Affinity Matrix

(F) Pearson vs DCCA

FIGURE 2.5: A comparison of how alternative methods compare to Pearson correlation in quantifying asset relationships between US stock returns. In all graphs Pearson correlation is on the x-axis. We can see how the different methods operate - for instance the partial correlation and dependency network methods tend to reduce correlations, showing how they are removing indirect relationships.

In this figure we can see the different types in action. Those which quantify non linear relationships (e.g. Kendall's τ/Mutual Information) tend to have 'fat' middles where assets can have significant values in one of the measures while an insignificant value in the other. This contrasts to the removal methods (Partial correlation, dependency partial correlation and affinity matrix) where the coefficient tends to have a lower value than with Pearson correlation, as they have removed some of the indirect relationships, while DCCA, a filtration method, is very similar to Pearson correlation, with the coefficients showing a similar pattern.

2.5 Random Matrix Theory

2.5.1 An Overview

Random Matrix Theory (RMT) is another potential solution for filtering correlation matrices to ensure only relevant information is left.

The idea behind this application of random matrix theory is that we compare the eigenvalues of the correlation matrix to those we would expect from a set of uncorrelated variables. A correlation matrix inferred from p uncorrelated Gaussian variables using n samples will have eigenvalues distributed according to the Marchenko–Pastur distribution.

$$p(\lambda) = \frac{n}{p}\frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\lambda} \tag{2.7}$$

Effectively this means that any eigenvalues that fall within the range λ_+ to λ_- are probably due to noise. λ_+ and λ_- are given as

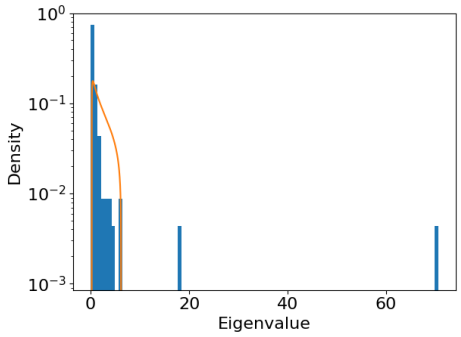
$$\lambda_\pm = (1 + \sqrt{\frac{p}{n}})^2 \tag{2.8}$$

Therefore we can remove any eigenvalues in this range from our dataset. Unfortunately this leaves us with the difficulty of what to replace them with. The general goal is to maintain the trace of the correlation matrix. To do so we can set the eigenvalues to 0, recreate the matrix and then insert 1s into the diagonal, this is the approach taken by [139]. Another approach is to take the mean of the noisy eigenvalues and replace these noisy eigenvalues with the mean.

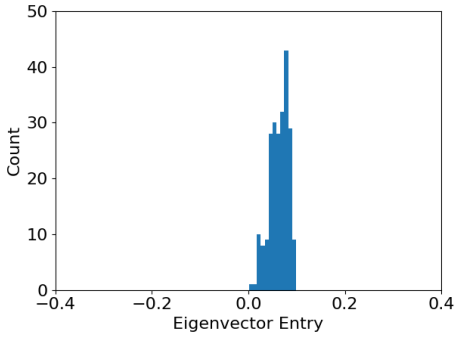
RMT also makes predictions about the entries of eigenvectors that correspond to these eigenvalues. Eigenvectors that are constructed from a correlation matrix inferred from uncorrelated Gaussian variables will have entries that are also Gaussian distributed, with zero mean. We show an example of the eigenvalue and eigenvector entry distributions from a correlation matrix constructed from the US stock returns (see section 2.2.8 for the data) in Figure 2.6.

2.5.2 RMT Analysis of Financial Markets

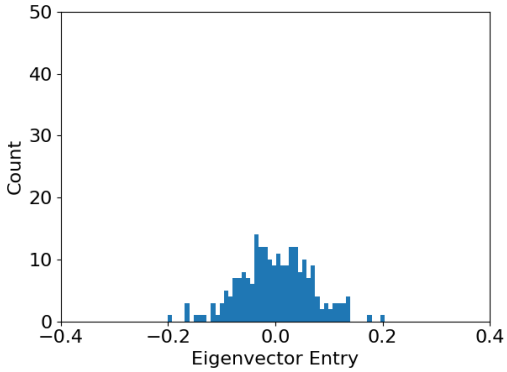
We can use RMT to answer the important question of whether correlation networks from asset returns contain real information or whether they are merely picking up noise present at that moment of time. Both Plerou et al. [206] and Laloux et al. [139] take this approach. Firstly, only a few of the eigenvalues and eigenvectors contained in the correlation matrix differ significantly from those obtained from a random matrix. This implies many of the relationships in the correlation matrix are noise. However the leading eigenvector (corresponding to the largest eigenvalue) tends to differ significantly and tends to contain the information that affects all stocks (for instance interest rate increases). Furthermore, the components of the leading eigenvector are significantly different from those obtained from a random matrix and have significant values for all stocks. The largest eigenvalue is also an order of magnitude larger than the next largest eigenvalue. They also study the next few eigenvectors corresponding to the next largest eigenvalues and find these tend to have significant values for related stocks - for instance those in the same sector or who have business in similar regions. Finally they show the leading eigenvector and corresponding eigenvalue tend to be preserved for correlation matrices inferred from data across different times, indicating they are stable over consecutive periods.



(A) Eigenvalue Distribution



(B) Leading Eigenvector Distribution



(C) 'Noisy' Eigenvector Distribution

FIGURE 2.6: Example distribution of eigenvalues from a stock return correlation matrix with the Marchenko–Pastur distribution highlighted (a), the distribution of entries from the leading eigenvector (the eigenvector that corresponds to the largest eigenvalue) (b) and the distribution of entries from an eigenvector associated with an eigenvalue lying within the Marchenko-Pastur distribution (c). In (a) we can see most of the eigenvalues lie within this distribution, but there are deviations, notably the largest eigenvalue. The leading eigenvector has entries that all are all positive, and approximately uniformly distributed (b), while the noisy eigenvector has entries than resemble a Gaussian distribution with a mean of 0 (c).

As previously mentioned, the largest eigenvalue of correlation matrices inferred from stock data is significantly larger than any other eigenvalue in the spectrum. Drozdz et al. [72] study why this might be the case. Showing that the distribution of off-diagonal entries in the correlation matrix follows a Gaussian distribution with non-zero mean and non-unit variance they propose a model of two matrices, one with entries drawn from a Gaussian distribution (G) and the other of all ones (U):

$$C = G + \gamma U \tag{2.9}$$

where $0 \leq \gamma \leq 1$. Furthermore they also show that the entropy in the system decreases during times of market turmoil and increases during times of market calm.

Borghesi et al. [31] further investigate the idea of removing the market mode from the correlation matrix and the effect this has on correlations at different time periods. They compare various different ways of removing the market mode - a one

factor model using the index, a one factor model using the average return, deleting the largest eigenvalue and eigenvector, and subtracting the average market return from the dataset. The one factor model using the average return has the greatest success in moving the mean correlation closest to zero, indicating this might be the best approach. They also note that removing the market mode from data of any return time period tends to give a similar distribution of correlation coefficients, indicating there is time invariant structure within the system.

Saeedian et al. [222] use RMT to study index returns from 40 countries, and attempt to quantify the amount of collective behaviour in this 'world stock market' using two measures they define based on the entries of the eigenvectors of the correlation matrix. The entries of the eigenvectors are less uniform than those from a randomly shuffled correlation matrix, indicating that there is a degree of collective behaviour in this world stock market. There is a strong degree of geographic clustering present in the dataset, and the Asian markets tend to behave more independently than the American or European markets. They also apply this model to stock returns from the US, UK, China and Iran to investigate their degree of collectiveness. The US scores the highest on the collectivity measure, followed by Iran, China and then the UK, indicating that more developed markets are not guaranteed to be more collective.

As previously mentioned, it is possible to uncover sector structure using RMT, with certain eigenvalues tending to have larger values for companies in the same sector. These values may not necessarily be positive though, with both large positive and negative values being present. Jiang and Zheng [117] study this for the Chinese and US markets, and in particular show that stocks which have large positive values in an eigenvector tend to have a much smaller average correlation with those that have a large negative sign, and visa versa.

Uechi et al. [266] extend RMT to a partial correlation network, as well as studying a normal correlation matrix. However their partial correlation network is only with respect to the market index and not a complete inversion of the correlation matrix. They then formulate a measure they term 'Sector dominance ratio' to study the influence of various sectors in different stock markets. They find that in the US and the UK, the financial sector is particularly dominant. This is also noticeable during the crisis of 2009, where the ratio actually goes negative for the sector. They also find that their partial correlation networks contain less noise based eigenvalues than the correlation networks.

2.5.3 Combining RMT and Filtration Methods

Once we have these filtered correlation matrices we can apply the same topological or threshold based models as above. The MST model is applied by Heimo et al. [103]. They find that the most central assets in the asset trees have the largest values in the leading eigenvector, and that companies in similar sectors are clustered together in said asset trees. They also state that this largest eigenvector has more equal entries than other eigenvectors of the correlation matrix.

The threshold model is applied by Namaki et al. [182]. To get a better idea of interactions in the system without the effect of the market mode, they construct a portfolio with the weights on each stock corresponding to the leading eigenvector and use a factor model to remove its effect.

$$M(t) = u_{\lambda_{\max}} \tag{2.10}$$

They model each stock return ($G_i(t)$) as a combination of the market effect, a constant and an error term:

$$G_i(t) = \alpha_i + \beta_i M(t) + \eta_i(t) \tag{2.11}$$

By performing least squares regression α_i and β_i can be estimated. Once this is achieved the residuals are calculated and then the network is constructed by using the correlations between the residuals. The networks are then thresholded at various values, and various network measures such as degree distribution and number of components are calculated.

2.5.4 Other Markets

Again in the case of random matrix theory analysis of the stock market, most of the attention has been on the US markets. Shen et al. [232] look at comparing the Chinese, US and Indian markets from a RMT perspective. Studying returns from the Shanghai stock exchange, and note that this stock market has a much larger mean correlation than that of the US, and in fact has no negative correlations at all. The largest eigenvalue is larger than the US or Indian stock markets, but there is less sector identifiable information in the eigenvectors than in the US markets, and in general less intra-sector correlation. Other authors have applied RMT analysis to the Indian [236] [136] and South African [282] markets.

If the reader is interested in more applications of random matrix theory in finance (for instance in portfolio optimization), we direct them to the following papers [32] [208].

2.6 Alternative Asset Applications

As well as stock markets, these models have also been applied to other types of financial markets.

2.6.1 Currency Returns

Minimum spanning trees have also been applied to currency returns. Currencies are however slightly more challenging to work with as there is no absolute value - only relative values compared to another currency. To compare currencies we can choose a single one as the base, and have all others priced relatively to this, but this influences the entire structure of the tree. Alternatively we can use pairs in the tree - i.e. we can use the change in USD-EUR and GBP-JPY rather than ensuring they are all priced identically. However the choice of pair matters (e.g. how many dollars buy a euro vs how many euros buy a dollar?) when using an MST model, as negative correlations are usually discarded by the method. McDonald et al. [168] propose to use all pairs (i.e. both USD - EUR and EUR to USD) to solve this, while Naylor et al. [183] uses only the value of said currency in USD and New Zealand Dollar (NZD) to reduce the complexity. These currency MSTs seem to show a strong degree of geographic clustering - countries that are close together in the world tend to also be close in the trees. The MST containing all pairs is also relatively stable, with around 50% of the edges in a tree maintained after a 2 year gap, while this stability is not studied for the USD or NZD based MSTs.

An alternative to a currency is to use a precious metal as the base for the MST. Mizuno et al. [176] construct MSTs using both the USD as the base and the price of an ounce of Platinum. In both MSTs there is clear geographic clustering, and in the

USD MST the metals form their own branch. In the Platinum MST the USD is the most central currency, with the Euro being the second most central.

Kwapein et al. [138] further exploring the effects of using a different currency base for these MSTs. Using gold as the basis of the first currency MST they find that the US dollar is by far the most central currency, followed by the Euro and the Singapore dollar. Using other currencies as the base gives very different results, indicating this choice is important to uncover structure. Looking at the MSTs from a Euro base, again the US dollar dominates as the most central currency, with a small amount of geographic clustering. If the US dollar is used as the base currency this gives a far less centralized tree, and branches of countries geographically close tend to form. Next the authors use a sliding window to study how the trees change over time. The MSTs that use GBP as the base currency are the most stable, followed by those than use Gold, then USD and finally the Euro. In general around 15% of the edges for the MSTs last more than 4 years.

Other authors have used the Real Exchange Rate (which is based on weighted basket of currencies) [198] [213], the Special Drawing Rights (SDR) [115] [275] and the Turkish Lira [129] as a base. Alternatively, dependency networks can be used to reduce the effects of currency choice [158].

Jang et al. [115] study how the currency MSTs vary over time. Using the SDR as a base currency, they construct MSTs from 61 countries using returns from January 1990 to December 2008, which covers many different financial crises. In contrast to stocks or indices, crises tend to cause a drop in correlation for currency returns, with the length of the MST usually increasing. Certain crises cause large changes in the MST structures, notably the Asian crisis of 1997. This crisis also starts the anticorrelation of the USD and the Euro, which has persisted since then. Even after these changes, the MSTs still show geographic clustering, with individual countries changing locations.

Vodenska et al. [269] apply a PMFG approach to stock index returns and to currency index returns from 56 countries. Dividing their dataset into 2 sections, a non-crisis period from 2002 - 2006 and to a crisis period from 2007 - 2012 they study both cross correlations and lagged correlations between stock indices and currencies, and perform community analysis using modularity maximization (the method of modularity maximization is unclear). Firstly they demonstrate that the mean correlation during crisis and non-crisis times is different and they confirm this using a t-test. They also show that the currency correlation coefficient distribution is significantly different from the stock index correlation coefficient distribution using a KS-test. From the stock index network, the community detection results mostly in geographically similar clusters on the non-crisis data. With the crisis data the clustering changes, notably with Portugal, Italy, Spain and Greece (sometimes referred to as 'PIGS') being clustered together, all of whom suffered from a sovereign debt crisis during this time. For currencies, the US dollar is placed in a community with many oil exporting nations. Many of these nations peg their currency to the US dollar, and in general the US dollar is used for trading oil, so we can see its influence here. The Euro is at the center of the European cluster. During the crisis period the European cluster is disrupted, perhaps due to the euro-zone crisis at this time.

2.6.2 Cryptocurrencies

Cryptocurrencies are another emerging field for analysis. Perhaps even more interestingly, they are not backed by anything evenly remotely real - there is no central bank or underlying asset that gives them value, making their return dynamics more exotic.

These are analysed using both the MST model and random matrix theory by Stosic et al. [245]. Interestingly they also look at partial correlation based versions of these MSTs. They find that the eigenvalue and eigenvector entry distributions are different to what random matrix theory would predict, and different to stock returns. In the MST, Bitcoin has the largest degree and most correlation, but not the largest betweenness centrality. This implies that it is influential in its community, but perhaps not the most influential in the entire market. The Bitcoin and Ethereum clusters are very close together, indicating these two are strongly correlated. In the partial correlation MST, the two are much further apart, and the overall tree has a significantly different structure. This implies that correlations between individual cryptocurrencies are strongly influenced by the entire market. Song et al. [239] propose to remove the influence of Bitcoin and Ethereum using linear regression, to allow the minor cryptocurrencies to form clusters. Using both an MST model and hierarchical clustering they identify six major clusters, where cryptocurrencies of similar types (e.g. providing anonymity, traded mostly in one country, token based) tend to make up these clusters. If the influence of Bitcoin and Ethereum is not removed then only two clusters form, as almost every other cryptocurrency is connected to one of the aforementioned two.

2.6.3 Index Returns

MSTs can also be applied to index returns [29]. This allows us to understand the interactions between countries as well as individual stocks. For instance Coelho et al. [57] study returns from the indices of 53 countries and create an MST from them. Generally countries that are geographically closer tend to be clustered together, with a developed European economy generally taking the most central node (usually France or Germany). Furthermore the trees have become shorter as time has gone on, implying countries are more connected. This also has implications for portfolio diversification - the more correlation that exists, the harder it is to construct portfolios from uncorrelated or anticorrelated assets. Partial correlation can also be used as a measure [277], which perhaps unsurprisingly gives different results. Here geographical clustering is less noticeable, and the US takes the spot as the most central node.

Sandoval et al. [223] further explore correlations between stock indices to see how countries are connected, and by extension, if geographic diversification can produce robustness to downturns. They study the effects of financial market crises on the correlations between international stock indices, showing that correlations also increase between them during times of market stress. This does unfortunately mean that it is challenging to diversify a portfolio, even if an investor looks to multiple countries. Correlations between indices also decay far more slowly than those between assets [273].

Another study on index returns in performed by Zhang et al. [290], who hope to understand the differences in MST structure between times of growth and times of recession. Focusing on the US economy, they chose to use index returns that represent various sectors of the US economy to get a coarse-grained view of the stock market as a whole. These index returns are clustered to separate them into times of market stress and market calm in a data driven way. Their conclusion is that the topology of the MSTs is significantly different during times of crisis, the MSTs are star like in times of market calm, with the industrial sector at the centre, but during times of crisis they become chain like.

Song et al. [238] study the returns from 57 market indices from countries across the world using a variety of time windows to show how different correlation dynamics are present at different time periods. Shorter time periods allow easier detection of

crises, with a 3 month window allowing the easy detection of the Asian 1997 crisis, Russian 1998 crisis and the 2008 global crisis, while longer time periods cause these to be smeared out. Constructing a PMFG using the returns, firstly they note there is a strong degree of geographic clustering, and this is verified using the InfoMap [220] community detection algorithm. The mean edge weight spikes during crises in these PMFGs as the correlation in the market increases. Mutual information is then used to measure edge changes between PMFGs adjacent in time. During the months in which shocks occur, the structure of the PMFG changes significantly, with drops then large spikes in the mutual information. Comparing the average correlation in the PMFG to that in an MST using Welch's t-test, they find the p-value spikes during times of stress, indicating that the two are detecting similar stress. Finally they study the eigenvectors and eigenvalues of the correlation matrices, focusing on the three largest eigenvalues and the eigenvectors that correspond to the two largest eigenvalues. As is usually the case, the largest eigenvalue is significantly larger than the second largest, and both it and the second largest tend to rise during times of market stress. The largest eigenvector tends to have positive entries for developed countries indicating there are common factors driving market growth, but small or negative entries for developing nations. The second largest eigenvalue has a more complex structure, with positive values for the US and some other European indices, but more negative values are present for some European, Asian and Oceanian indices.

As well as studying worldwide indices, authors have focused on studying the interplay between more specific indices, for instance in South Asia [22], the Industrials sector [101] or for Automotive Companies [133].

2.6.4 Bonds

Government bonds are a very commonly traded asset that makes up significant proportions of many portfolios, particularly as they are regarded as a low risk investment. Gilmore et al. [92] study minimum spanning trees constructed from monthly returns of bonds from 20 developed countries. As with the index return MSTs, there is a significant degree of geographic clustering, noticeably with the Eurozone being relatively tightly connected. Other countries that share strong economic links or cultural ties are also connected, for instance the UK and the US. Interestingly the mean correlation between the Eurozone countries has increased over the life of the dataset, while the mean correlation between other countries has decreased. This implies that the Eurozone countries have become more tightly linked, but also that there are greater opportunities for portfolio diversification.

Using this knowledge that the bond MSTs do reflect some real world economic characteristics, they can then be used to analyze various crises than occur. For instance Dias analyzes daily bond yields during the Eurozone sovereign debt crises over two papers [66] [67].

In the first paper [66] the period from April 2007 to October 2010 is studied for 19 EU countries. Four MSTs are constructed, one for the entire dataset, and then three for each third to study the dynamics over time. Over the entire dataset the countries most implicated in the crisis (Portugal, Ireland, Greece and Spain) form a branch in the tree, as do the Eastern European countries who joined the EU in 2004. Before the financial crisis the Eastern European countries still form a branch, and the Eurozone itself is tightly clustered. During the financial crisis the Eurozone cluster disintegrates, and Greece, Ireland and Portugal make up a branch. For the latter third of the dataset the PIGS cluster has formed, with Italy also near, and the

Eurozone cluster has mostly reformed. In general the topology of the MST does not change much for the entire dataset.

The second paper [67] covers the period from July 2009 - March 2012. In this paper the author focuses specifically on the Eurozone, and constructs both minimum spanning trees (i.e. connecting the most similar countries) and maximum spanning trees (i.e. connecting the most dissimilar countries) for each half of the dataset. For the maximum spanning tree, Germany and Greece and the central hubs with all other countries connected to one of them in both halves of the dataset. In the minimum spanning tree again there is a Portugal, Ireland, Spain and Greece cluster in both halves too. The author looks at quantifying the difference between the two trees by summing the edge weights of each tree and calculating the difference between these. This difference measure shows how the Eurozone bond yield synchronisation collapses during the crisis, as the different countries face different economic issues.

2.6.5 Commodities

Sieczka and Holyst [235] construct a minimum spanning tree from the returns of 35 commodities. Firstly they use RMT to analyze the eigenvalues and eigenvectors. They find the majority of eigenvalues lie outside of the Marchenko–Pastur distribution, implying there is a large amount of signal present. In their minimum spanning tree, metals tend to make up the most central nodes, with silver, copper and gold have the highest betweenness centrality. Furthermore they find the mean correlation between commodities increases as time goes on, particularly during the financial crisis.

Tabak et al. [248] take a similar approach, constructing an MST from the returns of 20 commodities. The commodities are divided into one of three sectors, Metals, Agriculture or Energy. They find a degree of clustering, with similar commodities making up branches in the MST. For instance, a branch of their tree is made up of Crude Oil, Unleaded Gas, Heating Oil and Natural Gas, all of whom are in the Energy sector. In general the Agricultural sector is the most influential and the Energy sector the least. However due to various oil price shocks throughout the period, the Energy sector has structure that varies the most.

Ji and Fan [116] concentrate on analyzing the correlations between oil index returns in 24 countries, with a mix of oil importers and oil exporters. As usual, there is a degree of geographic clustering involved in the MSTs, for instance the UK and Norway share an edge, as do the US and Mexico. However there are other political relationships also present, with the US and Iraq sharing an edge, and China being connected to the Middle East, due to Chinese imports of Middle Eastern oil. The US and Angola are the two most central countries in the MSTs. The high centrality of the US is easily explainable, being the largest oil producer in the world as well as a large importer, and the authors explain that Angola is not only a large exporter, but also produces oil with similar properties to the US. Studying the correlations over time, they find that the mean correlation is increasing, indicating markets are becoming more integrated and that the trees are relatively stable, with 64% of the edges persisting for over a year.

2.6.6 Other Assets

Returns are not the only data available about stocks, volatility is also an option. Micciché et al. [172] study minimum spanning trees inferred from both the correlations between stock returns and the correlations between stock volatilities. Those inferred from volatilities tend to be less stable than those inferred from returns, but the degree

structure from both tends to be stable over a number of years, evolving relatively slowly.

Real estate is another asset class which is a popular investment choice. Wang and Xie [274] study correlations between real estate indices from 20 countries using PMFGs and MSTs from 2006 - 2012. Most of the correlation coefficients from this time period are positive, and any negative correlations are small, indicating there is some synchronisation and influence between the markets. Surprisingly the US is not the most central country in these networks, instead France and the Netherlands occupy this role. In both the MSTs and the PMFGs there is strong geographic clustering, with branches and cliques tending to contain countries in similar regions. This is further verified using the Louvain community detection algorithm on the PMFG, although in this case the US and Canada join a European cluster rather than the Asian one. The mean correlation increases during the financial and sub prime mortgage crisis, and this corresponds to a decrease in the average distance in the networks. Looking at the single step survival ratio (i.e. fraction of edges that are maintained from one tree to the next), the MSTs and PMFGs tend to maintain most of the edges (around 95%) indicating the influences are stable in time. At least 80% of the edges are maintained for 20 days, although the life of an edge is dropping as time goes on, implying the correlation between markets might be decreasing. Finally the authors apply some RMT analysis, find that only three of the eigenvalues could be considered noise and that the eigenvector that corresponds to the second largest eigenvalue shows geographic localisation, with positive values for the Asian countries and negative values for the European countries.

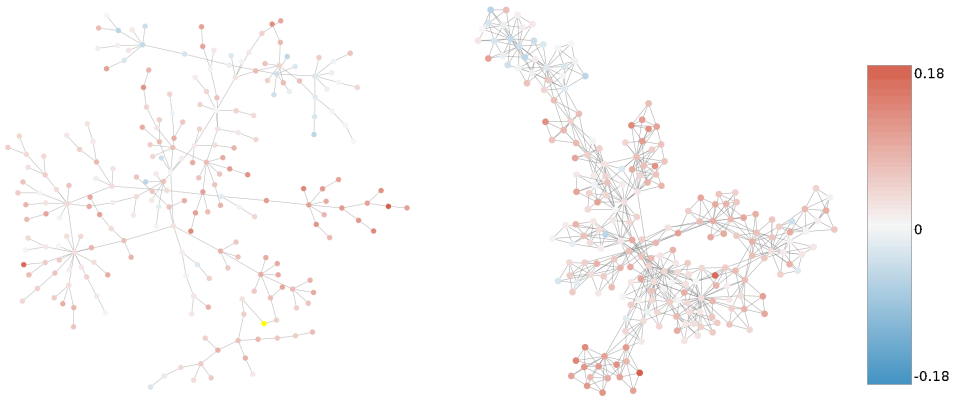
Meng et al. [171] apply both RMT and clustering based analysis to the US housing market. In this they find correlations between house price returns has increased sharply over time, particularly after 2002 (and note there was a significant house price bubble from 2002 - 2008 in the US), and when clustering the effect of geography is also weaker than perhaps would be expected.

2.7 Applications for Portfolio Selection

An obvious application of these networks is to make recommendations for portfolio selection. As previously mentioned, Onnela et al. [196] briefly point out that the companies which the minimum risk Markowitz portfolios put the highest weight on tend to be on the leaves of a minimum spanning tree.

Pozzi et al. [210] show that superior out of sample Sharpe ratios can be achieved by constructing portfolios from stocks on the periphery of these market graphs. The Sharpe ratio is defined as the return of a portfolio divided by its risk, and allows us to quantify both the risk and return of a portfolio. This is further studied by Peralta et al. [203] who point out the similarities between selecting stocks of a low centrality in a correlation network and the classic Markowitz minimum risk portfolios. It has however been pointed out that there is no strict algebraic relationship between graphs and Markowitz portfolios [111], and furthermore, other authors have actually found that there is positive correlation between the centrality of an asset and it's out of sample Sharpe ratio [123].

We show an example of this in Figure 2.7. Here we construct an MST and PMFG from the US stock return dataset from section 2.2.8, and then colour each node according to its out of sample Sharpe ratio on the following 504 days of data (from 2002-03-05 to 2004-03-02). In this specific example, there is relatively little



(A) MST (B) PMFG

FIGURE 2.7: Example graphs with the nodes coloured according to their out of sample Sharpe ratio on the next time window. Darker red means a more positive Sharpe ratio, while darker blue means a more negative Sharpe ratio.

relationship between the centrality and Sharpe ratio, and our work on the topic has also found relatively little relationship (see Chapters 5 and 6 for this)

A combination of RMT and topological filtering is used by Li et al. [150] to create optimized portfolios. They then compare just using the largest eigenvalue and corresponding eigenvector (i.e. the market mode) with using the full cross correlation matrix to create PMFGs. With these PMFGs they then create portfolios from peripheral stocks using a heuristic measure that combines several different forms of centrality. Comparing the peripheral portfolios to a equal $\frac{1}{n}$ portfolio on stock returns from the S&P500 and the Chinese CSI300, they find that the peripheral portfolios from the PMFG that only use the market mode out-perform both the naive and the full cross correlation PMFG. These methods do also generally select around 17 stocks continuously (out of a total portfolio of 20), meaning that the portfolios are relatively sparse and stable.

A different tack is taken by Boginski et al. [27], who use the market graph to attempt to create well diversified portfolios. Usually to create a market graph, we select edges with a weight above a set threshold. In this case they do the opposite, selecting edges with a weight below the threshold. They then look at forming portfolios from cliques in the graph. A clique is a set of nodes that are all connected to each other. In this situation a clique is a set of companies that are all uncorrelated with each other, which is ideal for diversifying a portfolio. However cliques are too restrictive in this situation, so instead they look at relaxing this to an *s-plex*. An *s-plex* is a subset of nodes in a graph where each node is only missing s edges to the others. Once they have inferred the graph, they set the weight of each edge in the graph to the return of the stock over that time period. They then look for the maximum-weight cliques and s-plexes in the graph - i.e. they want a set of uncorrelated stocks which have high returns. Testing their idea on windows of 250 days of return data from US markets they find they outperform both the S&P500 index and the Dow Jones, although it is unclear if this is out-of-sample data.

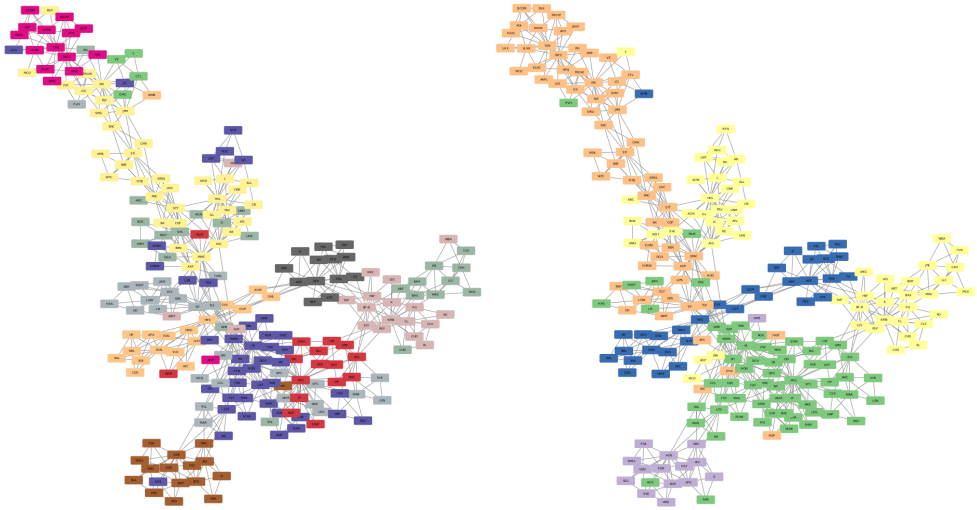
Motif structures in TMFGs are used by Turiel and Aste [264] to construct portfolios. A motif is a small recurrent subgraph that is statistically overrepresented in the network. Firstly they investigate the persistence of tetrahedral and triangular motifs in 4 markets, the US, Italy, Germany and Israel. Motifs tend to be maintained for longer periods of time (and hence more persistent) in the more mature and liquid

US market than the others. Two types of portfolio are constructed, one consistent of the ten most persistent triangular motifs and then a second type where each vertex is weighted by the inverse of motif persistence for the motifs it is present in. Perhaps unsurprisingly, the portfolios that consist of the most persistent motifs have a much larger out of sample risk compared to a random portfolio, while those that are weighted by the inverse of motif persistence have a much lower than expected out of sample risk.

Rather than taking a network approach, we can use the filtering methods to construct a correlation matrix with a reduced dimensionality and use this as input into a minimum risk portfolio. Tola et al. [256] use the standard MST and ALMST to filter the correlation matrices and compare them to RMT and unfiltered correlation matrices. If we assume that the future volatilities are known then the ALMST filtered correlation matrix generally achieves the best results out of all of the methods, achieving the lowest risk for a given return, while the standard MST correlation matrix does not out perform the unfiltered correlation matrix. However on a more realistic test, where the future volatilities were not known and short selling was forbidden, none of the methods showed consistently improved performance.

2.8 Clustering and Community Detection

It has been mentioned previously that intra-sector correlation is stronger than inter-sector correlation in these networks, and so it is only natural to see if we can extract this further using some form of community detection algorithm. We show an example of running the Louvain community detection algorithm [25] on a PMFG constructed from the same dataset as in section 2.2.8 in Figure 2.8. From this we can see there is a significant overlap between the sector assignments and the communities, showing it is possible to extract this information.



(A) Sector Membership Colouring (B) Community Membership Colouring

FIGURE 2.8: Example of running a community detection algorithm on a PMFG constructed from US stock returns. There is significant overlap between the communities detected and the underlying sectors in this example, showing that the sector structure can be extracted from the correlation networks.

2.8.1 Proposed Algorithms

Kocheturov et al. [134] use a clustering algorithm on correlation networks inferred using returns data from Sweden and the US. Obtaining the clusters by solving a P-median problem to infer star based communities, they find the clusters are actually more stable during times of market disruption. This implies that actually there is more noise during times of calm than market stress and that companies tend to fall and fail together.

Many authors tend to use the classic Newman formulation of modularity, and then some form of modularity maximisation algorithm to perform community detection in these correlation matrices. It has however been noted by MacMahon and Garlaschelli [157] that there are some issues with using Newman's modularity formulation. In this classic formulation of modularity the null model is that a community is no more connected than would be expected from the degree of the nodes involved. However this is not appropriate when we are looking at correlation matrices. In this situation a more appropriate null model is that of an uncorrelated system. The new function is therefore

$$Q = \frac{1}{m}\sum_i\sum_j (A_{ij} - I_{ij})\delta(c_i, c_j) \tag{2.12}$$

where I is the identity matrix, A is the adjacency matrix of the graph, c is a vector containing the community assignments and $\delta(c_i, c_j)$ is the Kronecker delta, equalling 1 if $c_i = c_j$ else equalling 0.

This formulation does deal with negative edges, but the cost is that modularity no longer goes from -1 to 1. The gain in modularity from moving isolated note i into community C can be calculated quickly and easily as follows

$$\delta Q = \frac{\sum_{j\in C} A_{ij}}{m} \tag{2.13}$$

Effectively this means we sum the correlation of node i with all the nodes in j. Furthermore they also propose a way in which the market mode can be removed in their method. Testing their method on S&P500 and FTSE100 constituent returns, they find it produces communities that are correlated within themselves and anti-correlated with the other communities. These communities are mostly are dominated by a single GICS sector, but no community is exclusively made up one of sector.

Teh et al [252] propose a very similar method to symbolisation to quantify relationships between financial assets. They term this method "digital correlation". It is defined as

$$DC(i,j) = \sum_{t=1}^{n} \theta(r_i(t), r_j(t)) \tag{2.14}$$

where $\theta(x) = 1$ if $x > 0$ else it equals 0. This removes the effects of outliers, and is quite similar to a rank correlation method. Secondly they propose a method "partial hierarchical clustering", which is based on the idea of creating one giant cluster from a seed node rather than joining individual clusters together. To discover the existence of smaller clusters within this model, they plot the complete linkage distance as more nodes are added to the cluster. When an elbow is reached (i.e. the distance increases significantly) this can be regarded as a smaller cluster. They then apply this method to the Singapore and Hong Kong stock exchanges, discovering that Chinese companies listed on these exchanges tend to form their own clusters, distinct from local ones.

2.8.2 Cluster and Community Analysis

A more detailed study into the clustering created from stock return correlations is conducted by Musmeci et al. [181]. In this paper they apply several linkage based, k-medoids and a novel method called the directed bubble hierarchical tree (DBHT) [241] to cluster unprocessed stock returns and those with the market mode removed. They use the adjusted rand index to compare the stock clustering to the underlying sector structure and a hypergeometric test to verify how much a sector is over-expressed in a given clustering. Firstly they find that removing the market mode improves the success of all the methods (something which has been known previously) in recovering the underlying sector structure. For instance the adjusted rand score of the DHBT method goes from 0.419 to 0.444 (the best performing on normal returns), and complete linkage increases from 0.387 to 0.510 (the best performing on detrended returns). Secondly the DHBT method shows the greatest over-expression of sector structure in the clusters produced when using a hypergeometric statistical test. Finally they show that for all methods the adjusted rand index drops during the 2007-2008 financial crisis.

There are multiple factors that contribute towards the clustering of a company, although by far the strongest of these is the sector that company is in. In [219] Ross uses a hierarchical clustering algorithm to cluster stock returns, firstly investigating the log returns of the S&P350 European stocks and then removing the sector specific returns from these stocks to obtain a more geographical based clustering. To evaluate the quality of their clustering they define a metric 'purity', which is designed for use in the hierarchical clustering algorithm. To calculate it, for each pair of companies in the same sector, we look at the point in the dendrogram where they are merged into the same cluster. At this point, we calculate the fraction of companies in this cluster that belong to the same sector as the two we have chosen. The mean across all pairs of companies in the same sector is taken, and this is used as the measure. With the log returns, the IT (0.93), energy (0.83) and Utilities (0.77) sectors have high purity, while the Consumer Discretionary (0.49) and Industrials have the lowest (0.33). If we remove the sector specific effects (by removing the average return in said sector) we get a strong increase in the geographic purity of the clustering - i.e. companies from the same country are more likely to be placed within the same cluster. Finally they study how the financial crisis effects the results, finding that the purity drops during the crisis but sharply rise afterwards.

The consequences for clustering using the eigenvalues and eigenvectors of the correlation matrix are studied by Heimo et al. [103]. While previous results have shown that sector structure is present in the entries of various eigenvectors [206] in this article the authors argue that many of the eigenvectors contain noisy entries, so distinguishing these noisy entries from signal is not trivial without prior knowledge (i.e. the actual sector structure). Furthermore they show that applying RMT to discard supposedly noisy eigenvalues does relatively little for increasing the quality of asset graph based clustering. Interestingly, removing noisy eigenvectors and eigenvalues has been studied in the bioinformatics domain and has been shown to increase the quality of clustering [20].

Other authors have used network based community detection methods. For instance Piccardi et al. [205] use the Newman formulation of modularity [184] and maximize it using the Louvain algorithm [25]. Analyzing both the DJIA (US) and MIB (Italian stock index) constituent company returns, they find the DJIA has a relatively high community structure, with 6 main communities which are mostly conserved over time. With the Italian market there is relatively little community

structure, possibly due to the common ownership of many Italian companies compared to the US. Isogai [113] uses the spectral strategy [184] for detecting communities in Japanese stock returns with the classic formulation of modularity. Performing statistical tests they show some clusters do reflect sector classifications, but others do not and they find the clusters can be used for portfolio diversification.

Another potential approach to understand how correlations change over time is to use multilayer networks. Rubin et al. [221] take this approach, inferring correlations from industry stock portfolios provided by Fama and French. One correlation matrix is inferred from a year of returns, so they end up with a network with 88 layers, ordered in terms of time. Performing community detection on this network using a Louvain-like method [25] which allows nodes to change communities through the layer they investigate how communities of companies in the US economy have changed over time. Before 1960 the communities are stable and do not change, but as we move forwards towards the present day the communities become less stable and the rate of change increases.

Filtered networks can also be used to perform community detection. This approach is performed by Nie et al. [187] using a PMFG. Using Newman's definition of modularity and spectral algorithm [184] they create communities from PMFGs inferred from the US and Chinese stock markets. They find that modularity firstly drops during the initial start of financial disruption, but then rises above previous values during these times, notably during the financial crisis of 2008.

Coronello et al. [60] study returns from the UK market for 5 minute and daily returns. Using an MST, PMFG, RMT and two forms of hierarchical clustering, average linkage and single linkage, they study sector clustering in the UK financial markets and how it differs over the two return time periods. All of the methods are capable of discovering sector clustering, and give different perspectives on the markets. For RMT, some of the non-noise eigenvectors tend to have larger entries for nodes that belong to one sector, however it is challenging to extract this information without first knowing the sector structure. The average linkage clustering algorithm has more success in recovering hierarchical structure than the single linkage method, and recovers more sub-sector clustering. Using the PMFGs they study how insular the sectors are, and find that the energy and financials sectors tend to form strong cliques with mostly intrasector edges. Comparing the clustering on 5 minute and daily returns, they find that the daily returns contain more sector clustering.

2.8.3 Applications

Applications of clustering include index tracking [71], where hierarchical clustering is used to cluster stocks with correlation as a similarity metric. We can then chose K clusters, and select 1 asset from each of these clusters. Theoretically this should be a diverse set of companies, and so we can then use these in a standard index tracking problem (i.e. regress these K companies against the index returns) to obtain the appropriate amount to invest in each stock. Purchasing every stock in the index is not desirable due to transaction costs, so this is an effective way of achieving a sparse portfolio. The authors claim their method is superior to randomly selecting stocks.

Papenbrock and Schwendner [199] take this approach, using k-means and hierarchical clustering on 125 correlation matrices inferred from assets in various classes, including indices, government bonds and commodities. Diving the correlation matrices into 5 classes, they then characterise these classes are bullish, bearish, uncorrelated and

volatile. After performing this they then cluster the asset returns, and study how these cluster structures change .

2.9 TopCorr

While most authors seem to develop custom code for each paper, there has been some software developed for the construction and analysis of correlation networks. Christensen et al [54] have created an R package, 'NetworkToolbox' that implements many of the methods we have mentioned, notably the PMFG, TMFG and the Dependency Network.

We have also developed a Python package on a similar theme, named TopCorr. This can be found at `https://github.com/shazzzm/topcorr`. Currently we have implemented

- TMFG
- PMFG
- MST
- kNN network
- Thresholding
- Dependency Network
- Partial Correlation
- Affinity Matrix
- Average Linkage MST
- MST Forest
- Detrended Cross Correlation Analysis

The package has relatively few requirements, needing only numpy, scipy and Networkx, and can be easily installed using pip.

2.10 Summary and Conclusion

In this chapter we have reviewed a substantial amount of the literature available on the inference of financial networks from asset return data. Often Pearson correlation is used to due its simple calculation and interpretation, but mutual information [80], rank correlation [209] [234], partial correlation [126] [277] [174], wavelet [276], affinity matrix [127] and DCC-GARCH [112] models have also been used in the literature.

The choice of which method to use to quantify the relationships does somewhat depend on the goals of the user. The standard Pearson correlation only detects linear relationships and is brittle to outliers and therefore might not be the best option. However, it is not easy to decide what is the best choice due to the noise, auto-correlations, non-linearities and outliers present in financial data. For instance, the affinity matrix method is good at picking up sector structure [127], but this means individual edges are less meaningful. Both the full partial correlation and

the dependency networks remove the market mode and provide an approach that is closer to a causal network than the standard correlation networks (note they are still not causal). The full partial correlation networks tend to be very unstable, so we would advise to not use them (see Chapter 5). The dependency networks are more stable however, and show more community structure [128] [126]. Mutual information can capture non-linear relationships that occur between asset returns [80] [81] [95]. However it has been shown that many of the differences between mutual information and correlation networks are due to univariate non-Gaussianity rather than non-linear relationships [100]. Both rank correlation and the DCC-GARCH models produce more stable networks than Pearson correlation [172] [112], at the cost of a less interpretable result.

Once inferred, we may wish to filter the networks due to the presence of a significant amount of noise. Topological forms of filtering, such as the MST [160] [196], PMFG [260] or TMFG [163] are forms that produce a graph with certain properties. Thresholding [26] [182] the correlation matrix is another alternative to this, but again there is no objective way of selecting said threshold, although bootstrapping [35] is a possibility [263] [61]. Alternatively a certain number of edges can be taken per node [187] but again this still requires a parameter to be selected.

These filtration methods have been run on more than just stock returns, with index returns [57] [223] [238], cryptocurrencies [239] [245], currencies [168] [183], commodities [248] [235] and bonds [92] [66] [67] among the options, as well as applications in non-financial fields [253] [254] [240]. When run on bonds, currencies or index returns, we generally see a large degree of geographic clustering [92] [66] , with countries that are close to each other tending to make up branches or communities in the resulting networks. Networks constructed from cryptocurrency returns are dominated by Bitcoin and Ethereum, which are by far the most popular cryptocurrencies [239] [245]. MSTs constructed from commodities have branches made up of similar commodities (for instance, crude oil, heating oil, natural gas) [248], with precious metals taking up the most central nodes [235].

The PMFG, MST and TMFG are guaranteed to produce a connected network, which is convenient for analysis. All three also produce networks with stable structures over time [172] [262] [264]. The PMFG is considered more stable than the MST [262]. This is probably due to it retaining more edges than the MST, meaning small changes in input are less likely to cause large changes in structure. However, both the MST and PMFG require large sample sizes to accurately recreate reference networks [121], which is concerning.

Furthermore, these topological filtration methods impose a certain structure on the data. It is not clear if this structure is correct or in general if there can be a correct structure for the markets. Previous work using MST filtered correlation matrices for portfolio optimization has not shown that they improve out of sample performance compared to using the full correlation matrices [256], and we feel that more work must be done to quantify whether these filtration methods actually produce improved results over the full correlation matrix in applications. We suggest that future work must involve investigating the performances of the methods with an objective measure. For instance, this could involve comparing the success of graph classification for filtered and full correlation matrices on financial returns (i.e. to distinguish between different market states), on gene expression data (e.g. to distinguish between controls and patients with cancer) and on MRI data.

Random matrix theory is another solution to filtering the correlation matrix, based on the idea of studying the eigenvalues and eigenvectors of the correlation matrix. The null model is that of uncorrelated Gaussian white noise, and we remove

any eigenvalues and eigenvectors that correspond to this null model. In general most of the eigenvalue/eigenvector pairs do correspond to Gaussian white noise, and the removal of these is shown to improve Markowitz portfolio performance [45]. In correlation matrices inferred from financial returns, the largest eigenvalue is significantly larger than the others in the matrix and its corresponding eigenvector gives us the effect the overall market has on the stocks. This drives most of the correlations in the market, with most of the companies being more correlated to the index than to anything else [230]. This effect is referred to as the market mode. If we remove this market mode, we get a correlation matrix with more scale invariant information [31]. We can of course then apply topological [103] or thresholding [182] based models.

The entries of the eigenvectors can be used for the detection of specific sectors, with the companies belonging to said sector having larger values in a particular eigenvector. However, this structure is difficult to pull out without prior knowledge [103].

Stock returns are not stationary. The usual solution to this problem is to select a small window to infer the correlation (or desired similarity quantification method) matrix. Selecting this window is challenging, as there is no objective way to do so, and the choice can have large effects on the resulting networks [291] [238]. Potential solutions to this problem include observation weighting [209] or change point detection [14]. Both of these methods still require hyperparameter selection.

The structure of the networks also varies over time. Times of market stress bring greater correlation [211], cause the MSTs inferred from them to shrink [196] [145] and by some measures have a larger amount of signal [72]. Interestingly it is also claimed the markets become more stable during these times, after an initial amount of disruption [125], with the clustering also becoming more stable [181] [134] although the success of obtaining the underlying sector structure decreases. Other authors have found the reverse (i.e. the networks become more unstable) [189] [188] [102]. This also seems to depend on which filtration model is used (if any at all). Correlations between indices from different countries also increase [223] [273], indicating that these effects are not geographically limited. The number of edges maintained between adjacent PMFGs can be used to predict the volatility of the next window [181].

In general companies in the same sector appear to share greater correlation than those in different sectors, and therefore tend to be connected in both MSTs and PMFGs to companies in the same sector [261] [194] [261]. These effects can be seen in intraday data, with the sector structure for certain sectors (e.g. financials, energy) emerging in even 5 minute returns [261]. It is generally accepted that the financial, industrial and consumer discretionary sectors tend to have a disproportionately large centrality in networks inferred from US stock data [126] [160] [266] [65]. For other countries, this can vary. The financial sector in the UK also shows a high level of centrality, but for Germany and Japan the Industrial sector is much more central, indicating these methods can detect differences in macroeconomic structure [266]. The different sectors also show differing levels of interconnectedness, with the financial sector having a large number of connections to both itself and other sectors, while the energy and technology sectors tend to only connect to other companies in the same sector [126] [65]. Developing markets tend to have higher correlations and less sector structure than developed ones [236]. The different filtration methods also tend to agree on sector centrality for the US markets [65].

We can attempt to recover these sectors by either running clustering algorithms using the distance metric (2.3) via various methods [134] [181] or by treating the

correlation matrix as a network using one of the many community detection algorithms available [157] [113] [205]. Removing the market mode also tends to improve the success of detecting the underlying sector structure [157] [181]. It is claimed this is due to the stocks mostly being correlated with the overall market, removing this dominating effect allows the intra-sector relationships to be detected more easily [230]. These clusters can then be used to create diversified portfolios [71]. The two most popular methods for sector detection are agglomerative hierarchical clustering and the Louvain method [25] in community detection, but the method that has the greatest success in recovering the underlying sector structure is the Directed Bubble Hierarchical Tree [181] [241]. If indices or stocks from different countries are clustered, geography tends to play the strongest role in deciding these clusters [219] [269]. During times of market stress, it becomes more difficult to detect the underlying cluster structure [181] [199]. There are disagreements on the stability of the clustering structure during these times of stress, with authors both finding it increases [134] and decreases [181] [179].

Investigation into using these networks for portfolio optimization has resulting in disagreements on the use of centrality for portfolio selection. In general, lower risk stocks are placed upon the peripheries of a network [194] [203]. However, some authors have found that better Sharpe ratios can be found by selecting these peripheral assets [210], while others have recommended selecting central assets in the networks [123]. There is also limited theoretical support for either observation [111].

An alternative approach is to use cliques to obtain supposedly unrelated stocks [27], clustering [71], motif participation [264] or we can use the topological filtering tools [150]. For these alternative approaches, the authors use a variety of baselines for comparison (some use the index return, some use a minimum risk portfolio and others use volatility weighted portfolio (i.e. investing in every asset proportionally to its volatility)) and do not use the same datasets. This makes it difficult to directly compare the proposed methods, or provide a recommendation on which performs the best.

In conclusion these correlation networks can reveal structure in the stock markets and help us to understand which companies influence the overall economy. However, many of the current approaches tend to take models that are convenient for the analyst rather than necessarily correct, and have few objective results on their performance. This is the case for both the methods which quantify relationships between assets, and the methods which filter these relationships to remove noise. Furthermore there are parameters to select in most cases (for instance, the window size), and few objective ways to select this parameter. These parameters also have large effects on the results, and authors tend not to share their code or data, making reproducibility challenging. For these networks to become useful to financial practitioners, these issues must be solved. Our suggestion for future work is to test these methods on both synthetic data with a known ground truth network, and on real data with an objective measure (for instance, graph classification) in order to provide recommendations on which methods have the best performance, and potential ways of parameter selection. Authors must also share their code and data to ensure it is reproducible. We suggest our software package, TopCorr may assist in this manner. Sharing of code and data may also assist in resolving some of the disagreements over the properties of the networks.

If the reader is further interested in the clustering of financial returns, we direct them to a review by Marti et al. [162] which is on a similar theme to this one.

Software

The code to produce the plots in this chapter was written in Python 3, and we make use of NumPy and SciPy [193] for handling the matrices, pandas [169] for handing the data, statsmodels [227] for calculating the Spearman and Kendall's τ correlation, scikit-learn for calculating mutual information [200], matplotlib [110] for plotting, Networkx [96] for the network analysis, TopCorr for the implementation of the methods and Cytoscape [229] for the graph visualization.

Chapter 3

A Comparison of the Performance of a Selection of Precision Matrix Estimators

3.1 Introduction

In Chapter 2 we reviewed the various methods of inferring a financial network. In this chapter we focus specifically on reviewing methods that can be used to infer a sparse precision matrix.

The precision matrix has various useful applications, such as in classification, portfolio optimization and in graphical modelling (for instance if i and j have a zero entry in their entry in the precision matrix, then they are conditionally independent if the data is Gaussian). It is also required to calculate the partial correlation matrix, which we utilize in Chapters 5 and 6 The precision matrix can be obtained by inverting the sample covariance matrix. However, if we have more dimensions than samples, which is often the case in finance and other fields, this sample covariance matrix is not invertible. Alternatively we may desire the precision matrix to have some property (or have some prior information to indicate this is the case). In the machine learning field regularization is a classic solution to both these problems.

When using these precision matrices for network inference, it is desirable that they are sparse. This makes the resulting network much easier to interpret. Therefore, firstly in this chapter we review a selection of methods which can be used to infer a sparse precision matrix.

There is also relatively little information of how these methods perform in uncovering the underlying model, both relative to each other and to a simple baseline. Therefore in section 3.4 we provide a software package which implements a selection of the methods mentioned in the previous chapter. Using this software, in section 3.5 we compare the performance of the estimation methods to each other and to a simple baseline of thresholding the covariance matrix.

3.2 Estimation of Sparse Precision Matrices

3.2.1 Neighbourhood Selection

To start with we introduce how to recover a precision matrix from a dataset using regression. Consider a p dimensional multivariate normal distribution, written as

$$X = \begin{bmatrix} \boldsymbol{x_1} \\ \boldsymbol{x_2} \\ \vdots \\ \boldsymbol{x_p} \end{bmatrix} \sim N(\boldsymbol{\mu}, \Sigma) \tag{3.1}$$

where $\boldsymbol{\mu}$ is the mean vector and Σ the covariance matrix. For simplicity we often assume the mean is 0.

Firstly we set up a least squares regression problem regressing one variable ($\boldsymbol{x_i}$) on the rest (X_{-i})

$$\arg\min_{\beta_i} \frac{1}{2n} ||\boldsymbol{x_i} - X_{-i}\boldsymbol{\beta_i}||_2^2 \tag{3.2}$$

where n is the number of samples. Each value of $\boldsymbol{\beta}$ allows us to gain row i of the precision matrix (Θ) by using

$$\beta_{ij} = -\frac{\theta_{ij}}{\theta_{ii}} \tag{3.3}$$

(this will be derived below) Doing this for every variable allows us to reconstruct said precision matrix. To show why this is the case, we construct the equation for the inverse of the covariance matrix $\Sigma\Theta = I$ into components and permute so i is the final variable

$$\begin{pmatrix} \Sigma_{11} & \boldsymbol{\sigma_{12}} \\ \boldsymbol{\sigma_{12}}^T & \sigma_{22} \end{pmatrix} \begin{pmatrix} \Theta_{11} & \boldsymbol{\theta_{12}} \\ \boldsymbol{\theta_{12}}^T & \theta_{22} \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \tag{3.4}$$

We can multiply these out to link the various values. We are interested in

$$\Sigma_{11}\boldsymbol{\theta_{12}} + \boldsymbol{\sigma_{12}}\theta_{22} = \mathbf{0}. \tag{3.5}$$

Our solution for $\boldsymbol{\beta_i}$ is

$$\boldsymbol{\beta_i} = \frac{1}{n}\boldsymbol{x_i}^T X_{-i}(X_{-i}^T X_{-i})^{-1} = \boldsymbol{\sigma_{12}}^T \Sigma_{11}^{-1}. \tag{3.6}$$

From 3.5 we can extract an equation for $\boldsymbol{\theta_{12}}$

$$\boldsymbol{\theta_{12}} = -\Sigma_{11}^{-1}\boldsymbol{\sigma_{12}}\theta_{22} = -\boldsymbol{\beta_i}^T\theta_{22}. \tag{3.7}$$

This is equivalent to (3.3). If we wish to estimate the diagonal we also must calculate θ_{22}. We show this using (3.5) and

$$\boldsymbol{\sigma_{12}}^T\boldsymbol{\theta_{12}} + \sigma_{22}\theta_{22} = 1 \tag{3.8}$$

we have that

$$\boldsymbol{\theta_{12}} = -\Sigma_{11}^{-1}\boldsymbol{\sigma_{12}}\theta_{22}. \tag{3.9}$$

Plugging this into (3.8)

$$-\boldsymbol{\sigma_{12}}^T\Sigma_{11}^{-1}\boldsymbol{\sigma_{12}}\theta_{22} + w_{22}\theta_{22} = 1 \tag{3.10}$$

$$\theta_{22}(\sigma_{22} - \boldsymbol{\sigma_{12}}^T \Sigma_{11}^{-1} \boldsymbol{\sigma_{12}}) = 1 \tag{3.11}$$

$$\theta_{22} = \frac{1}{\sigma_{22} - \boldsymbol{\sigma_{12}}^T \Sigma_{11}^{-1} \boldsymbol{\sigma_{12}}} = \frac{1}{\sigma_{22} - \boldsymbol{\sigma_{12}}^T \boldsymbol{\beta_i}} \tag{3.12}$$

which allows us to calculate it from $\boldsymbol{\beta_i}$. It is also equal to the inverse of the residual of problem (3.2). To show this we firstly multiply out the L_2 norm of (3.2)

$$\boldsymbol{x_i}^T \boldsymbol{x_i} - 2\boldsymbol{x_i}^T X_{-i} \boldsymbol{\beta_i} + \boldsymbol{\beta_i}^T X_{-i}^T X_{-i} \boldsymbol{\beta_i} \tag{3.13}$$

$$\sigma_{22} - 2\boldsymbol{\sigma_{12}}^T \boldsymbol{\beta_i} + \boldsymbol{\beta_i}^T \Sigma_{11} \boldsymbol{\beta_i} \tag{3.14}$$

Substituting in $\boldsymbol{\beta_i} = -\frac{\boldsymbol{\theta_{12}}}{\theta_{22}}$

$$\sigma_{22} + 2\boldsymbol{\sigma_{12}}^T \frac{\boldsymbol{\theta_{12}}}{\theta_{22}} + \frac{\boldsymbol{\theta_{12}}^T}{\theta_{22}} \Sigma_{11} \frac{\boldsymbol{\theta_{12}}}{\theta_{22}} \tag{3.15}$$

Now using (3.9) we can reduce the quadratic term

$$\frac{\boldsymbol{\theta_{12}}^T}{\theta_{22}} \Sigma_{11} \frac{\boldsymbol{\theta_{12}}}{\theta_{22}} = -\frac{\boldsymbol{\theta_{12}}^T}{\theta_{22}} \boldsymbol{\sigma_{12}} \tag{3.16}$$

which turns (3.15) into

$$\sigma_{22} + 2\frac{\boldsymbol{\sigma_{12}}^T \boldsymbol{\theta_{12}}}{\theta_{22}} - \frac{\boldsymbol{\theta_{12}}^T \boldsymbol{\sigma_{12}}}{\theta_{22}} = \sigma_{22} + \frac{\boldsymbol{\sigma_{12}}^T \boldsymbol{\theta_{12}}}{\theta_{22}} = \sigma_{22} - \sigma_{12}^T \beta_i \tag{3.17}$$

which we can see is similar to (3.12)

Once we have estimated the precision matrix (Θ), a simple re-scaling will give us the partial correlation matrix (P)

$$P_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} \tag{3.18}$$

From this we can gain some intuition as to the meaning of partial correlation. It's definition is "the correlation between two variables once the linear effects of all others have been removed". Looking at equations (3.2) and (3.3) we get that the partial correlation between i and j is proportional to the weight of predicting variable i from variable j (and visa-versa as it is symmetric).

For the applications we are interested in, this method has some issues. It will return a dense result (i.e. every entry of $\boldsymbol{\beta}$ will have a non-zero value) and that it may not work when we have more dimensions than samples ($p > n$). Since we are interested in inferring a network, we do not want there to be an edge between every node, and in finance we are often interested in datasets where we have more dimensions than samples. Therefore we turn to regularization to solve this problem. Meinshausen and Buhlmann [170] use a L_1 lasso penalty on a regression problem to estimate the edges of a node using the measurements from a multi-dimensional system

$$\arg\min_{\beta_i} \frac{1}{2n} ||\boldsymbol{x_i} - X_{-i}\boldsymbol{\beta_i}||_2^2 + \lambda ||\boldsymbol{\beta_i}||_1 \tag{3.19}$$

where the L_1 penalty is the sum of the absolute values of β

$$||\boldsymbol{\beta}||_1 = \sum_{i=1}^{p} |\beta_i| \tag{3.20}$$

The L_1 norm is known for giving sparse results [255], and will give us good results even when we have more dimensions than samples.

3.2.2 Graphical Lasso

The precision matrix produced by neighbourhood selection is not guaranteed to be positive semi-definite or symmetric. This is an issue in many applications, for instance, in classification and portfolio optimization. To achieve these goals an alternative approach is required. Friedman et al [86] propose a coordinate descent algorithm to solve the L_1 penalized Gaussian log-likelihood equation, defined as

$$\arg\min_{\Theta} -\log|\Theta| + tr(\Sigma\Theta) + \lambda||\Theta||_1 \tag{3.21}$$

We can rewrite the optimization problem (3.19) as

$$\arg\min_{\beta_i} \frac{1}{2}\boldsymbol{\beta_i}^T\Sigma_{11}\boldsymbol{\beta_i} + \boldsymbol{\beta_i}^T\boldsymbol{\sigma_{12}} + \lambda||\boldsymbol{\beta_i}||_1 \tag{3.22}$$

and instead of estimating the rows of the precision matrix separately, we iteratively update Σ_{11} with its new row after we have estimated the corresponding row of the precision matrix. To demonstrate the similarity between (3.21) and (3.22) we look at the gradient equation for 3.21, calculated by differentiating it with respect to Θ

$$\Theta^{-1} - \Sigma - \lambda.\text{Sign}(\Theta) = \mathbf{0} \tag{3.23}$$

If we use our partitioned matrices (defined in (3.4), focus on the top right hand corner and let $W = \Theta^{-1}$) the equation becomes

$$\boldsymbol{w_{12}} - \boldsymbol{\sigma_{12}} - \boldsymbol{\gamma_{12}} = 0 \tag{3.24}$$

Again using (3.4) we can see that

$$\boldsymbol{w_{12}} = -W_{11}\frac{\theta_{12}}{\theta_{22}} \tag{3.25}$$

$$\boldsymbol{w_{12}} = W_{11}\beta \tag{3.26}$$

where $\beta = -\frac{\boldsymbol{\theta_{12}}}{\theta_{22}}$ Now our gradient equation is

$$W_{11}\boldsymbol{\beta} - \boldsymbol{\sigma_{12}} - \boldsymbol{\gamma_{12}} = 0 \tag{3.27}$$

This is the stationarity condition for the optimization problem (3.22) where $\Sigma_{11} = W_{11}$. Therefore by solving this problem we can recover $\boldsymbol{w_{12}}$ and $\boldsymbol{\theta_{12}}$ - i.e. we can recover both the precision and the covariance matrices without having to invert either. The solution produced is also guaranteed to be positive semi-definite and symmetric. To show this, we refer back to the matrix partitions (3.4), and note we are updating one column of the matrix. Using the Schur complements, the determinant

of the matrix is

$$\det \Theta = \det \Theta_{11} \det(\theta_{22} - \theta_{12}^T \Theta_{11}^{-1} \theta_{12}) \tag{3.28}$$

For each iteration of the algorithm, we are only changing θ_{12}. If we assume the input into each stage (Θ_{11}) is positive definite, this means that we require

$$\det(\theta_{22} - \theta_{12}^T \Theta_{11}^{-1} \theta_{12}) > 0 \tag{3.29}$$

for the resulting matrix Θ to be positive definite. The inside of the determinant is a scalar, so we can discard the determinant

$$\theta_{22} - \theta_{12}^T \Theta_{11}^{-1} \theta_{12} > 0 \tag{3.30}$$

This becomes

$$\frac{1}{\sigma_{22}} > 0 \tag{3.31}$$

which is always the case. It is also common to add λ to the diagonal of the sample covariance matrix before inputting it into the algorithm as a form of regularization, and this will further guarantee that the output is always positive semi-definite.

There are some tricks that can be used to speed the graphical lasso up. For instance, if we can threshold and permute the sample covariance matrix at a threshold λ such that it becomes block diagonal, the precision matrix produced by the graphical lasso will also be block diagonal [166]. This further means that we can threshold the sample covariance matrix (an operation of trivial complexity) to extract the block diagonals, and then run the graphical lasso individually on each block. Since the complexity of the glasso is around $O(p^3)$ this provides a significant increase in performance. Furthermore there are situations in which the graphical lasso and thresholding the sample covariance matrix perform the same [237]. QUIC [106] uses iterative quadratic approximation based on Newton's method to solve much larger problems than the coordinate descent solution (i.e. 1000 - 200,000 variables) , and BigQUIC [107] increases the performance even more (200,000 - 1,000,000 variables). Being able to deal with such large systems allows us to not have to pre screen variables before inferring a network - for instance the application the authors suggest is fMRI data, where they apply it on a dataset containing 228,483 dimensions.

However in our experiments, and noted by Tan et al [250] the graphical lasso tends to create isolated nodes with no connections and long 'tails', where nodes at either end of the tail are dissimilar. In their paper, they explain this that the way the glasso determines which components are connected is equivalent to single linkage clustering. To fix this they propose to use a different method to detect clusters and then running the graphical lasso on each cluster individually.

3.2.3 Other Approaches

Using the formulation of the problem 3.19, we can use any kind of regression method to obtain a precision matrix. Peng at al [202] propose SPACE (Sparse PArtial Correlation Estimation) on this theme. This a form of neighbourhood selection, but instead of solving the problem row by row, we solve one large problem over the entire graph. Firstly we note that we can express the regression parameter in terms of the partial correlation coefficients and the diagonal of the precision matrix

$$\beta_{ij} = P_{ij} \sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}} \tag{3.32}$$

Therefore we can impose sparsity on the partial correlation matrix rather than the precision matrix using this link, which avoids imposing the uniform structure on the network, helping to improve the detection of hubs. The optimization problem requiring solving is the following

$$L(P,\Theta) = \frac{1}{2}\sum_{i=1}^{p} ||\boldsymbol{x_i} - \sum_{j\neq i}^{p} P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}||_2^2 + \lambda \sum_{1\leq 1<j\leq p} |P_{ij}| \tag{3.33}$$

We alternate between solving for P and for the diagonal of Θ until the problem converges - this typically takes around two to three iterations. To solve this we formulate a large lasso problem - on the order of np by $p(p-1)/2$. This is done by creating two new regression variables $\hat{Y}$ and $\hat{X}$. $\hat{Y}$ is a np column vector containing all the variables stacked upon each other

$$\hat{Y} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_p})^T \tag{3.34}$$

Next we need to create a matrix that contains the predictors for the appropriate $\hat{Y}$ values.

$$\hat{X} = (\bar{X}_{(1,2)}, \dots, \bar{X}_{(p-1,p)}) \tag{3.35}$$

$$\bar{X}_{(i,j)} = (0,\dots,0,\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}^T,0,\dots,\sqrt{\frac{\Theta_{ii}}{\Theta_{jj}}}\boldsymbol{x_i}^T,0,\dots,0) \tag{3.36}$$

The following problem is then solved to calculate the off-diagonal values of the partial correlation matrix

$$\min_{\boldsymbol{a}} ||\hat{Y} - \hat{X}\boldsymbol{a}|| + \lambda||\boldsymbol{a}||_1 \tag{3.37}$$

where $\boldsymbol{a}$ contains the off-diagonal values. Once we have run this we can estimate the diagonal as following

$$\frac{1}{\Theta_{ii}} = \frac{1}{n}||\boldsymbol{x_i} - \sum_{j\neq i}^{p} P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}||_2^2 \tag{3.38}$$

Another regression and variable selection procedure is the Dantzig selector, proposed by Candes and Tao [50]. It is formulated as follows

$$\begin{aligned} \underset{\beta}{\text{minimize}} \quad & ||\boldsymbol{\beta}||_1 \\ \text{subject to} \quad & ||X^T(\boldsymbol{y} - X\boldsymbol{\beta})||_\infty < \lambda \end{aligned} \tag{3.39}$$

Instead of searching for the solution with the lowest error given a constraint on the weights as before, we are looking at finding the 'simplest' solution for a given error bound. This selector has some attractive theoretical properties, but its predictive capabilities are generally not as good as the lasso [73] and the L_∞ norm can make it more susceptible to outliers than a standard least squares problem.

A Dantzig selector inspired procedure is CLIME (Constrained L_1-minimization for Inverse Matrix Estimation), proposed by Cai et al [47]. They propose to solve the

following problem

$$\begin{aligned} \underset{\Theta}{\text{minimize}} \quad & ||\Theta||_1 \\ \text{subject to} \quad & ||\Sigma\Theta - I||_\infty < \lambda \end{aligned} \tag{3.40}$$

Again this is solved in a column by column way, formulating p optimization problems. The problem for the ith column is

$$\begin{aligned} \underset{\Theta}{\text{minimize}} \quad & ||\boldsymbol{\beta}||_1 \\ \text{subject to} \quad & ||\Sigma\boldsymbol{\beta} - \boldsymbol{e_i}||_\infty < \lambda \end{aligned} \tag{3.41}$$

where e_i is a p dimensional vector with a 1 in the ith position and zeros elsewhere.

Sparse Column-wise inverse operator (SCIO), developed by Liu and Luo [153], is a method inspired by CLIME that formulates a quadratic problem to infer the matrix. Here they propose a column-wise loss function

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta} - \boldsymbol{e_i}^T\boldsymbol{\beta} \tag{3.42}$$

where e_i is a vector of zeros with a 1 in the ith position and Σ is the covariance matrix. If we differentiate this we get

$$\Sigma\boldsymbol{\beta} - \boldsymbol{e_i} = 0 \tag{3.43}$$

And rearranging we get

$$\boldsymbol{\beta} = \boldsymbol{\theta_i} \tag{3.44}$$

where $\boldsymbol{\theta_i}$ is the ith column of the precision matrix. Therefore we can see that solving this optimization problem for $i = 1...p$ dimensions will give us a precision matrix. However this problem is neither sparse nor solvable when $p > n$. Therefore they add a L_1 penalty term to solve this

$$\min_{\boldsymbol{\beta}} \frac{1}{2}\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta} - \boldsymbol{e_i}^T\boldsymbol{\beta} + \lambda||\boldsymbol{\beta}||_1 \tag{3.45}$$

SCIO tends to give sparser results than the glasso or CLIME, usually with better performance (time wise) than either (this can be up to several orders of magnitude) as we can use the fast coordinate-descent based methods to solve it rather than using linear programming. We perform some analysis of this later on in this chapter (see Table 3.1). The authors also provide some theoretical analysis of using cross-validation as an approach to decide upon a regularization parameter.

All of these methods do require the selection of a regularization parameter. While there are methods to do so (e.g. BIC, cross validation e.t.c) some authors have worked on methods that perform automatic parameter selection. We explain the basic idea here. To recover the true non-zero coefficients using the lasso, we require that the threshold is set high enough to ensure that the weights assigned to a variable due to noise are set to 0. Effectively this means setting $\lambda = \sigma\sqrt{\log p}$ [50] [49]. This requires estimating the noise (σ), which can be done so if we have the oracle solution ($\boldsymbol{\beta}^*$) of the problem

$$\sigma^2 = \frac{1}{n}||\boldsymbol{y} - X\boldsymbol{\beta}^*||_2^2 \tag{3.46}$$

Unfortunately we do not have the oracle solution. An approach to deciding the regularization parameter is the "Scaled Lasso" [246]. With the scaled lasso we estimate the lasso problem and then the noise in the problem iteratively, updating

the regularization parameter until convergence. Starting we a guess for λ (i.e. solve the lasso problem for a chosen λ), we estimate the noise, use this to rescale our λ to a more correct value and repeat until it converges. This proceeds as follows:

- $\sigma = 5$, $\lambda_0 = \sqrt{2p/n}$, $\lambda = \sigma\lambda_0$
- $\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \frac{1}{n}||\boldsymbol{y} - X\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1$
- $\sigma^2 = \frac{1}{n}||\boldsymbol{y} - X\hat{\boldsymbol{\beta}}||_2^2$
- Repeat until the difference in σs is less than 0.0001

Once we have this solution, we can then simply perform the row by row regression to calculate each row of the precision matrix. This does not require any special software beyond the usual machine learning libraries, which is attractive.

Other options on a similar front use the SQRT-Lasso [19] [152], but the challenge with this method is actually solving this problem practically - it may be convex but cannot be easily entered into many existing optimization packages.

3.2.4 Ledoit-Wolf Covariance

All the above methods impose sparsity to regularize. This may not always be desirable, and so next we review a method that produce dense precision matrices.

Ledoit and Wolf [142] propose a shrinkage method to create a regularised covariance matrix. By proxy this also gives us a well regularized precision matrix. Their model combines the identity matrix and the sample covariance matrix to produce a new estimate as follows

$$S_* = (1 - \lambda)S + \lambda\frac{tr(S)}{p}I \tag{3.47}$$

To decide λ the authors wish to minimize the Frobenius norm of the difference between Σ_{lw} and the true population covariance matrix Σ_*

$$\min_{\lambda} E[||\Sigma_* - \Sigma_{\text{lw}}||_F^2] \tag{3.48}$$

The optimal solution of ρ is

$$\rho = \frac{E[||S - \Sigma_*||_F^2]}{E[||S - tr(S)I||_F^2]} = \frac{\beta^2}{\delta^2} \tag{3.49}$$

The interpretation here is that if S is very close to Σ_* (i.e. our estimate of the covariance is good) then we do not need to shrink much, or if our shrinkage choice is not very good then we should not shrink much either. However the obvious flaw so far is that we need to know the true population covariance matrix to obtain the correct value for ρ - and if we did then we would not need to bother estimating it to begin with! We therefore require estimates of β^2 and δ^2. We can estimate δ^2 as following:

$$\hat{\delta^2} = ||S - tr(S)I||_F^2 \tag{3.50}$$

and β^2 as

$$\hat{\gamma^2} = \frac{1}{n^2}\sum_{k=1}^{n}||x_k x_k^T - S||_F^2 \tag{3.51}$$

$$\hat{\beta^2} = \min(\hat{\delta^2}, \hat{\gamma}) \tag{3.52}$$

$$\hat{\rho} = \frac{\hat{\beta}^2}{\hat{\delta}^2} \tag{3.53}$$

The constraint on $\hat{\beta}^2$ ensures that $\rho < 1$. While it is rarely necessary, it does help stop us accidentally making our estimate less well formed. This is of importance in the application they are interested in - portfolio optimization, which requires a positive semidefinite matrix for the optimization procedure to be convex (this is further explored in Chapter 4).

Slightly tighter bounds can be achieved if the data is assumed to be Gaussian [53] or alternatively cross validation can be used to select λ [280].

3.3 Parameter Selection

Most of these methods require a regularization parameter to be selected to ensure the correct level of regularization. This is a hard problem that has not been decisively solved. The most common choice is to use cross validation, where we divide our data into a training set and a test set, fit the model with various parameters on the training set and use a loss function to evaluate the model on the test set. The usual choice of loss function is the Gaussian log-likehood function, but other choices could include the Frobenius norm or the quadratic loss ($\text{Tr}(\Sigma\Theta - I)^2$)). Once we have a parameter we then refit the model on the entire dataset. For certain methods (e.g. SCIO) a per column parameter can also be selected. This can give better results, but also can result in overfitting, particularly when the number of samples is small.

Cross validation does not take into account the complexity of the model produced. We can use the Bayesian information criterion (BIC) [226] or the Akaike criterion (AIC) [2] to fix this. For LASSO problems, they are defined as [295]

$$\text{BIC(M)} = -2 * L + \ln(n)\text{ed}(M)$$

$$\text{AIC(M)} = -2 * L + \text{ed}(M)$$

where L is the log likelihood, given by equation 3.21, and ed(M) is the effective model dimension, in this case given by the number of edges in the network (the number of non-zero values from the off-diagonal values of matrix M) These can be minimized to find an optimal value of λ.

There are of course also the methods mentioned above to automatically calculate the appropriate value of λ from the noise in the problem (Scaled Lasso and SQRT-Lasso)

3.4 NITK

The Network Inference Toolkit (NITK) is a software package written in Python. Its goal is to provide an sklearn compatible implementation of the following algorithms

- Neighbourhood Selection (NS)
- Scaled Lasso
- SPACE
- CLIME
- SCIO

- Thresholding the Covariance matrix (Threshold)

While most of these estimators have been implemented in certain author provided scripts, to the best of our knowledge they have not been implemented in one complete package. The estimators all have unit tests so their results can be compared to the original author's implementations to ensure they are correct. This software can be acquired at https://github.com/shazzzm/nitk.

We show some example code below to estimate a precision matrix using the Scaled Lasso below:

```
from nitk import scaled_lasso
import numpy as np
from sklearn.datasets import make_sparse_spd_matrix

p = 50
n = 20
K = make_sparse_spd_matrix(p)
C = np.linalg.inv(K)
X = np.random.multivariate_normal(np.zeros(p), C, n)
sli = scaled_lasso.ScaledLassoInference()
sli.fit(X)
print(sli.precision_)
```

3.5 Comparing the Methods

3.5.1 Approach

Using NITK we compare the various precision matrix estimation methods on graphs with several different degree structures. To create the precision matrices we use the normalized Laplacian (i.e. the off-diagonal row entries normalized to add to 1, and the diagonal set to 1) of the graph with the desired structure. This is then combined with the identity matrix to ensure it is positive definite as follows

$$\Theta = \alpha L + (1 - \alpha) I \tag{3.54}$$

α is usually set to 0.8.

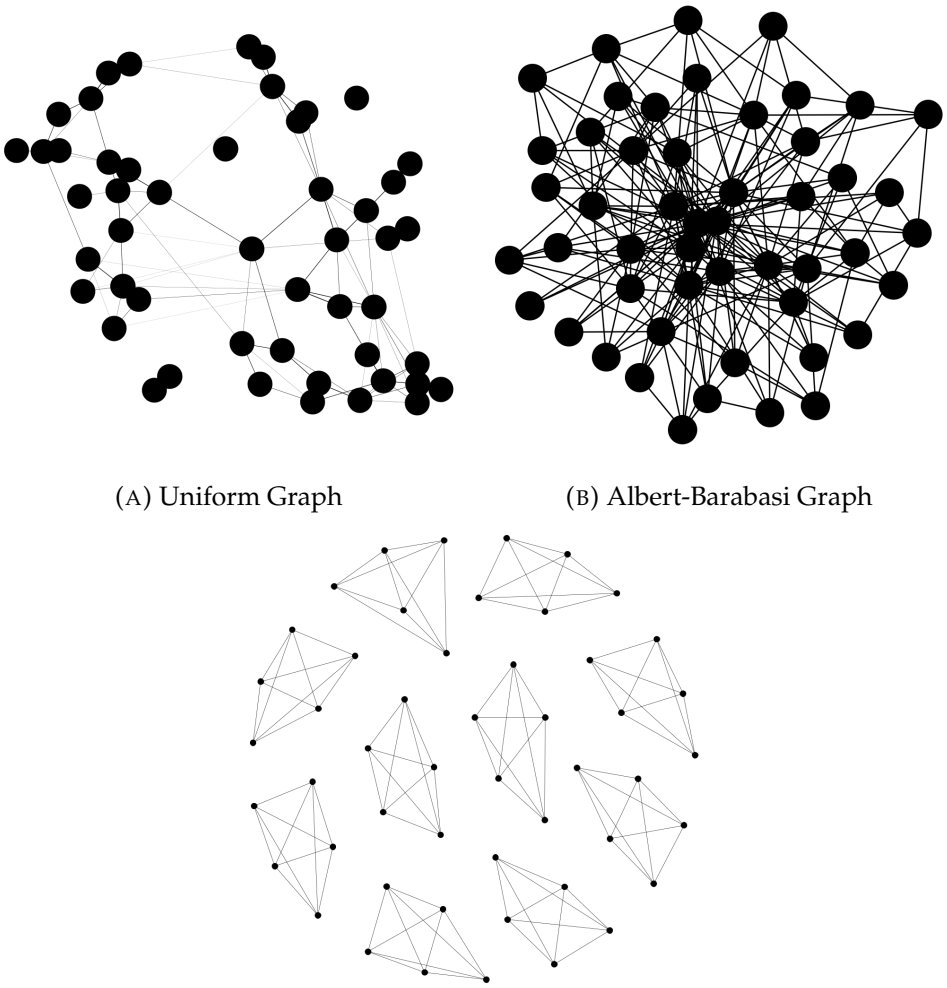
We construct networks with 3 different degree structures. The first structure, has a relatively uniform distribution, and is constructed using the *make_sparse_spd_structure* method from scikit-learn.

The second has a power law degree distribution and is constructed using the Albert Barabasi model [12]. Here, we start with an empty graph. New nodes are then added until we have the desired number. Each new node is connected to 5 existing nodes using preferential attachment (i.e. new nodes are more likely to be connected to those of higher degree.)

The third is a modular structure, constructed using a caveman model. These networks consist of fully connected modules with no connections to other modules. Each module is composed of 5 fully connected nodes.

Example networks with these structure are shown in Figure 3.1. We choose these models as they are likely to cover many of the networks that are likely to be seen in real life networks [12] [202].

To evaluate the success of the methods we use several measures. Firstly, we infer precision matrices for a set of regularization parameters. We then calculate



(A) Uniform Graph

(B) Albert-Barabasi Graph

(C) Caveman Graph

FIGURE 3.1: Example networks for the chosen structures. Each network has 50 nodes

the true positive rate and the false positive rate for each regularization parameter in order to create a ROC curve. An ROC curve is a graph with the false positive rate on the x-axis and the true positive rate on the y-axis. It allows us to quantify the performance of a binary classifier when the threshold for classification is varied.

For our experiments, we calculate the area under the ROC curve using the trapezium rule. This quantifies how well the method does in uncovering the true model without having to select a regularization parameter.

Often with applications of these inference methods we are not just interested if it is possible to pick up the underlying model, but specifically if we can select a singular network to analyze. We also explore the success of the methods in picking up a specific network. The optimal parameter is selected using 5 fold cross validation, and we use the F_1 measure of their success in uncovering the underlying sparsity structure. The loss function for the cross validation is the log likelihood.

These different network structures are generated for five combinations of n and p. These combinations are $n = 50$ $p = 50$, $n = 50$ $= 100$, $n = 50$ $p = 200$, $n = 100$ $p = 200$ and $n = 200$ $p = 500$. For each given p and n we generate 20 different precision matrices, generate a dataset from this precision matrix, run an estimation procedure on this dataset and evaluate the mean and standard deviation of the various measures mentioned. The set of regularization parameters is decided firstly by calculating the smallest parameter for which the networks will be completely sparse ($\lambda_{\max}$). We then take the minimum regularization parameter as $10^{-3}\lambda_{\max}$, and use log spaced values from the minimum to the maximum λ.

We compare all the methods implemented in NITK bar CLIME. In initial experiments, CLIME had a high level of success in uncovering the underlying structure, but simply took too long to be practically useful for anything other than trivial problems, as it scales poorly with the number of dimensions (p).

3.5.2 ROC Curves

Firstly we compare the success of the methods using the area under the ROC curve (AUC). We show the results in Figures 3.2 (uniform), 3.3 (power law) and 3.4 (caveman).

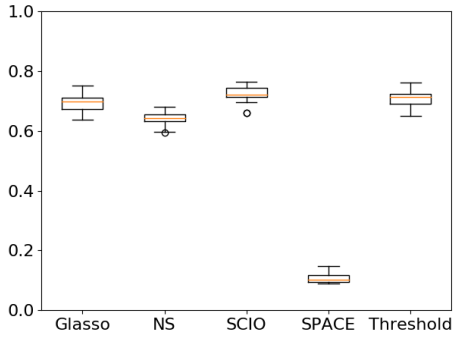
We see similar results across the different network structures. Firstly we note that thresholding performs surprisingly well, generally being competitive with the best performing sparse precision matrix estimator across all p and n combinations and the different network structures. For the caveman networks, the threshold estimator seems to perform particularly well.

For the sparse precision matrix estimators we notice a few trends. SCIO performs consistently well across all the combinations of structure and p/n combinations. SPACE tends to perform better as n and p get larger, which the authors also find in their original paper [202]. Neighbourhood Selection (NS) shows the opposite effect, performing significantly worse as p and n get larger. With the exception of the uniform networks, the performance of the glasso is consistent, being the second or third best estimation method.

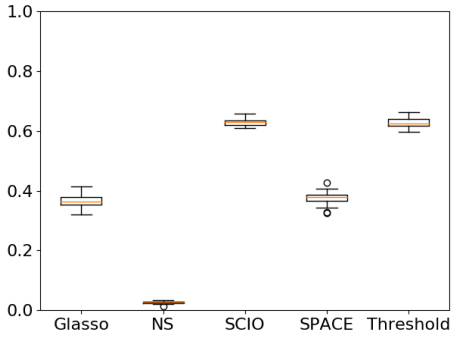
We also find here that the network structure tends to influence the success of uncovering the underlying network, with the caveman structure generally having the highest mean AUC, and the power law the lowest.

3.5.3 F_1 Scores

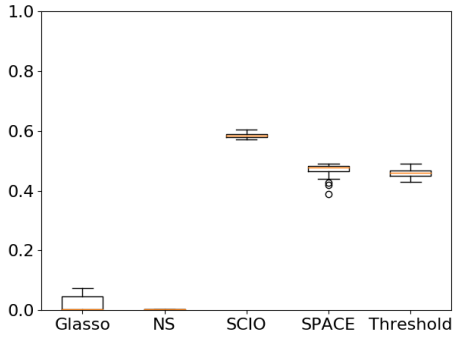
As previously mentioned, we are also interested in selecting up one particular network for many applications. In this section we explore the success of the various methods



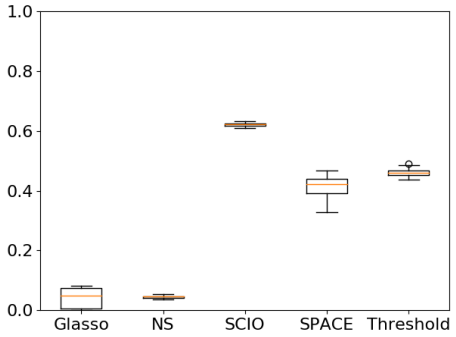
(A) $n = 50, p = 50$



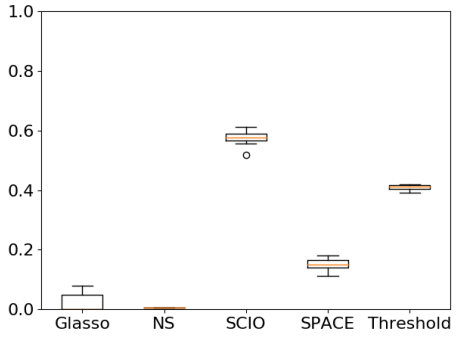
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$

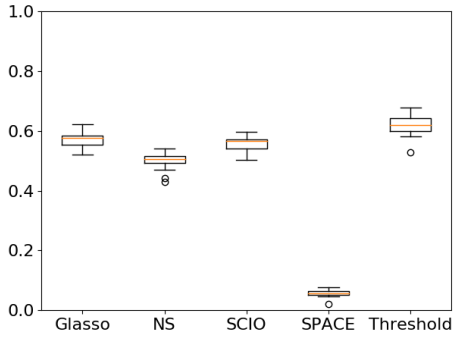


(D) $n = 100, p = 200$

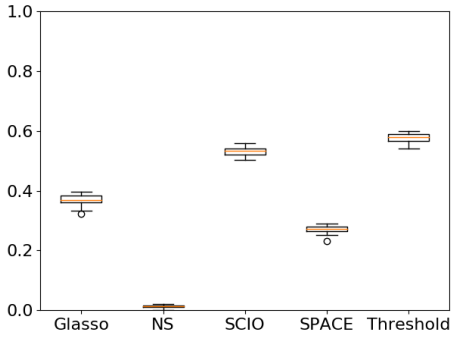


(E) $n = 200, p = 500$

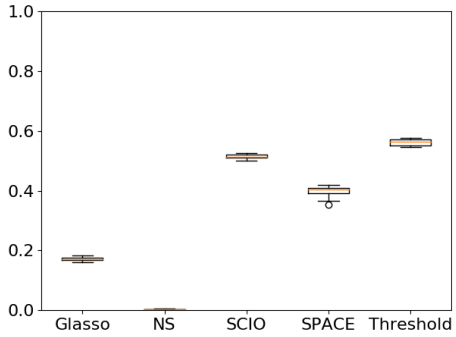
FIGURE 3.2: AUC for graphs with an underlying uniform structure



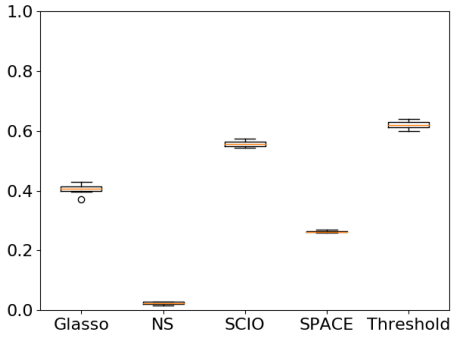
(A) $n = 50, p = 50$



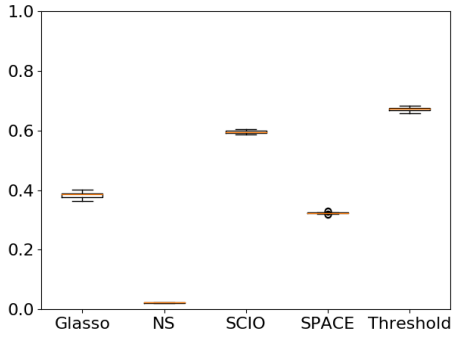
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$

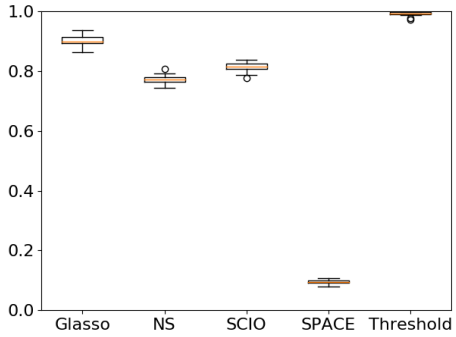


(D) $n = 100, p = 200$

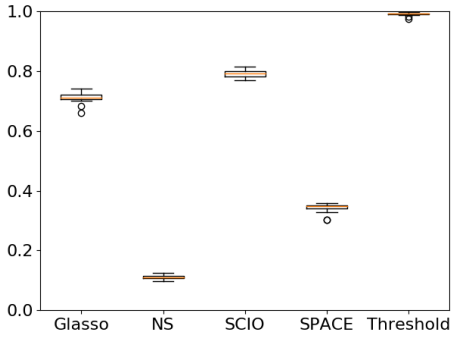


(E) $n = 200, p = 500$

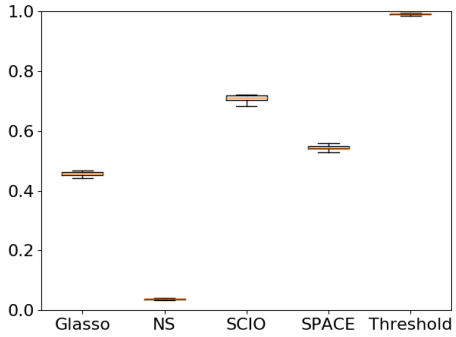
FIGURE 3.3: AUC for graphs with an underlying power law structure



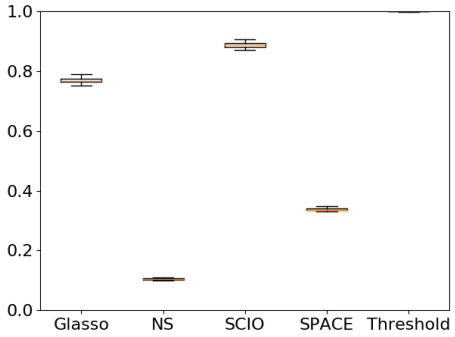
(A) $n = 50, p = 50$



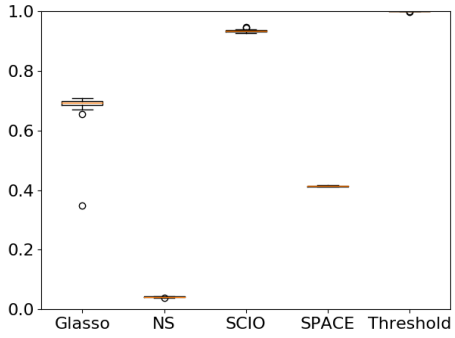
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$



(D) $n = 100, p = 200$



(E) $n = 200, p = 500$

FIGURE 3.4: AUC for graphs with an underlying caveman structure

in picking up the locations of the non-zero values when we select one particular network from the set inferred. To select this network we use 5-fold cross validation. The results are shown in Figures 3.5 (uniform networks), 3.6 (power law) and 3.7 (caveman).

Starting with the uniform structure, we note that the performance of any method is not particularly high, with none achieving an F_1 score of above 0.4. The performance of the methods also depends on p and n, with the Scaled Lasso, the glasso and neighbourhood selection performing better when p is smaller and n larger, while SPACE again tends to perform better when p and n are both larger. The threshold method is relatively consistent across all tested values of p and n. SCIO tends to fail as p increases.

Next we look at the results for the networks with a power law structure. For almost all of the methods, the results are worse compared to the uniform structure. Thresholding appears to be the best performing method here, followed by SPACE. All methods also have a worse performance as p increases.

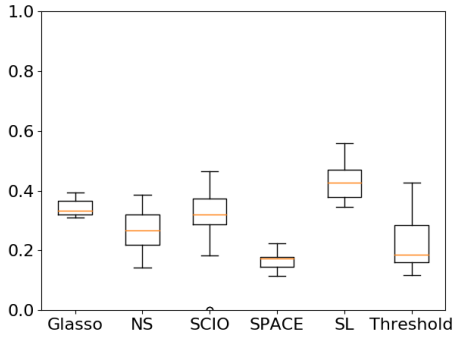
The results for the networks with a caveman structure are very different to the others, and quite surprising. Firstly we note that most of the methods perform much better than on the other two network structures, aside from SPACE and the threshold, with most of the methods achieving an F_1 score of above 0.4 on every combination (compared to achieving an F_1 score below 0.4 on the uniform networks) We also see different results when p and n are changed, with Neighbourhood Selection and the Scaled Lasso increasing in success as p and n are increased, while thresholding and SPACE actually decrease. Unlike before, the glasso seems relatively unaffected by different values of n and p.

For the uniform and power law networks, the relative performance of the methods seems quite similar (i.e. if a method performs well compared to the others with AUC as a measure, it also performs well compared to others with F_1 score as a measure). However, this is not the case with the caveman networks, with the threshold method in particular performing poorly.

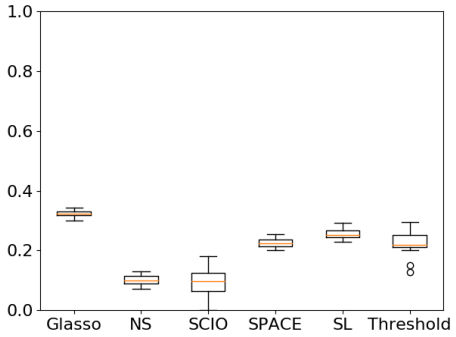
3.5.4 Performance

Finally we compare the performances of the various network inference methods. This is more complex than would be expected due to the dependence of the performance of the lasso on the regularization parameter chosen - it will take longer to converge for a smaller regularization parameter (and therefore a less sparse result). Furthermore the scaled lasso does not have a parameter to chose - making it difficult to achieve a comparable result. Therefore we decide to measure performance for a variety of parameters, in a way that would replicate attempting to select a regularization parameter using an information criterion. For the methods where it is possible, we calculate regularization parameter at which the output matrix will contain only zeros on the off-diagonal. This is our maximum regularization parameter. We then divide this by 1000 and use this as our minimum regularization parameter. Using these we create a set of log spaced values from the minimum to maximum regularization parameters and run the estimation procedure for each value (as proposed in the glmnet package [84]), timing how long it takes in total. We run these tests 50 times and measure the mean and standard deviation. This is only run for the uniform degree distribution. We do not show results for thresholding as the cost is trivial. A table of these results is shown in Table 3.1.

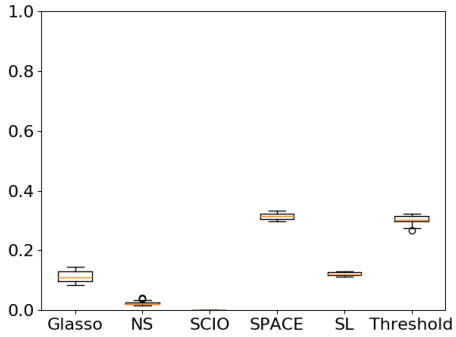
The impressive performance of SCIO is notable, with it generally being the quickest out of the methods that have to check a variety of regularization parameters. The



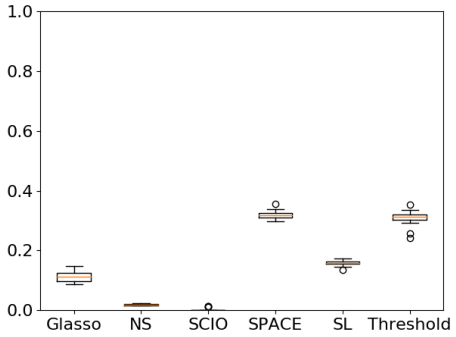
(A) $n = 50, p = 50$



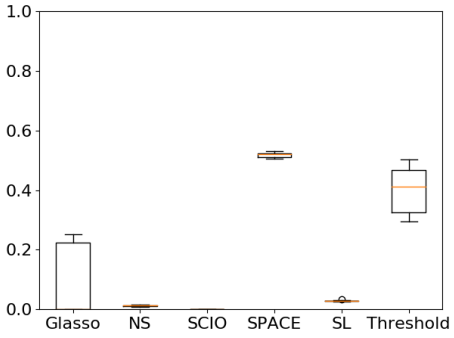
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$

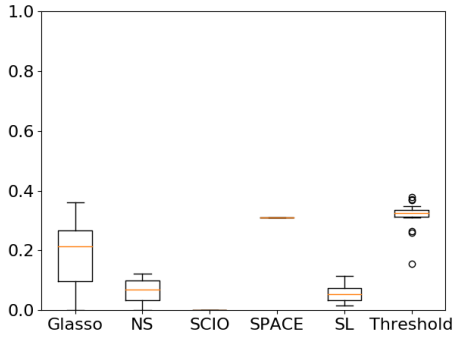


(D) $n = 100, p = 200$

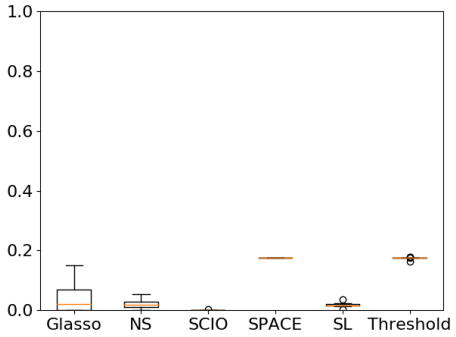


(E) $n = 200, p = 500$

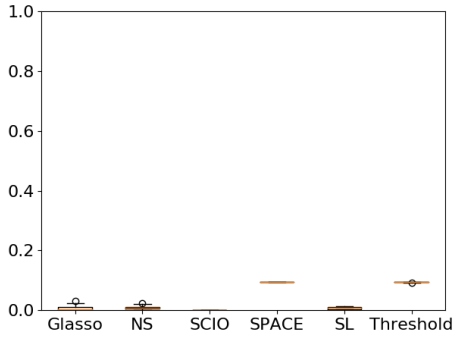
FIGURE 3.5: F_1 score for graphs with an underlying uniform structure



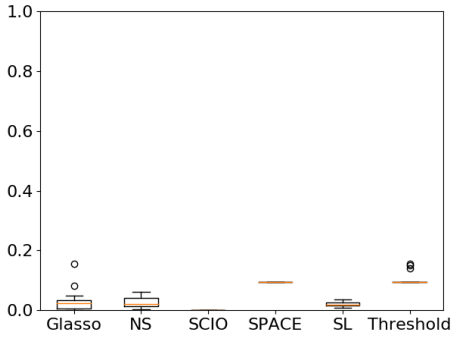
(A) $n = 50, p = 50$



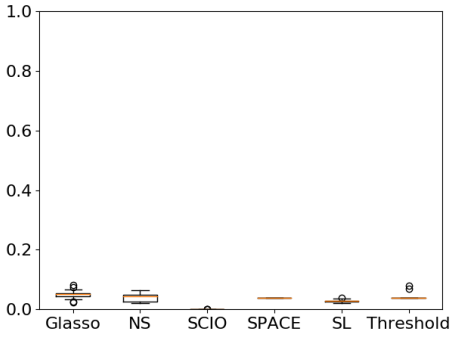
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$

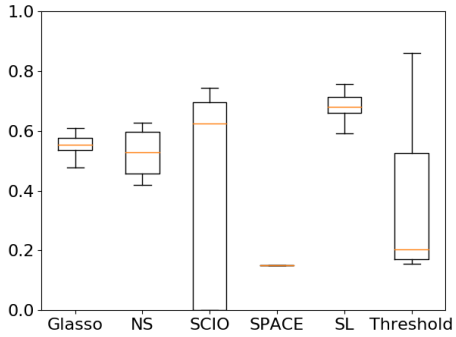


(D) $n = 100, p = 200$

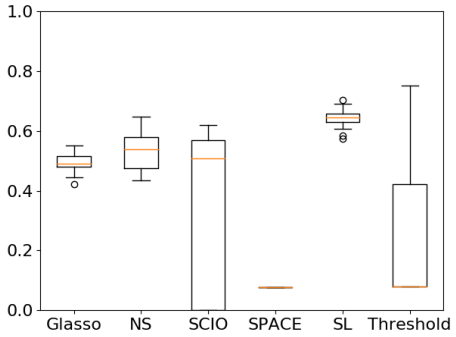


(E) $n = 200, p = 500$

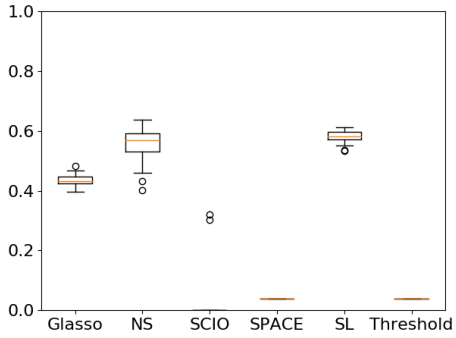
FIGURE 3.6: F_1 score for graphs with an underlying power law structure



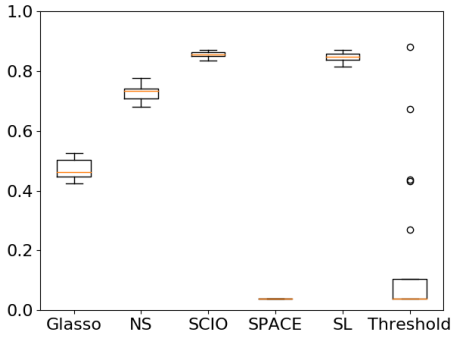
(A) $n = 50, p = 50$



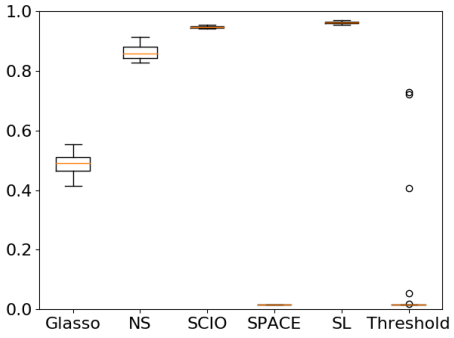
(B) $n = 50, p = 100$



(C) $n = 50, p = 200$



(D) $n = 100, p = 200$



(E) $n = 200, p = 500$

FIGURE 3.7: F_1 score for graphs with an underlying caveman structure

Method	p	n	Uniform
Graphical Lasso	50	10	4.291 ± 1.423
NS	50	10	2.837 ± 0.197
SCIO	50	10	0.743 ± 0.010
Scaled Lasso	50	10	0.079 ± 0.021
SPACE	50	10	1.350 ± 0.120
Graphical Lasso	100	10	10.436 ± 3.061
NS	100	10	7.483 ± 0.694
SCIO	100	10	3.163 ± 0.133
Scaled Lasso	100	10	0.141 ± 0.019
SPACE	100	10	6.785 ± 0.499
Graphical Lasso	100	50	23.924 ± 4.530
NS	100	50	11.266 ± 2.946
SCIO	100	50	3.849 ± 0.366
Scaled Lasso	100	50	0.442 ± 0.042
SPACE	100	50	52.343 ± 2.207
Graphical Lasso	200	50	28.915 ± 11.497
NS	200	50	24.039 ± 3.733
SCIO	200	50	17.560 ± 1.038
Scaled Lasso	200	50	2.725 ± 0.653
SPACE	200	50	383.200 ± 43.622
Graphical Lasso	500	50	92.974 ± 40.034
NS	500	50	46.320 ± 13.517
SCIO	500	50	134.983 ± 9.858
Scaled Lasso	500	50	69.990 ± 29.665
SPACE	500	50	4056.933 ± 126.811

TABLE 3.1: Mean and standard deviation of the time taken for the algorithms to estimate a precision matrix for 50 different regularization parameters. This is then run 50 times. The advantages of not having to select a regularization parameter (the Scaled Lasso calculates one from the data) are clear for performance purposes.

advantages of automatically selecting a regularization parameter are obvious with the Scaled Lasso - even though it must iterate and solve a succession of lasso problems, it still is far more efficient than even the quickest method that requires parameter selection.

3.6 Conclusion

In this chapter we have provided a review of a set of methods to infer a sparse precision matrix, described a software package which implements them, and compared how those methods perform to each other, and to a baseline of thresholding the sample covariance matrix.

Most of the work in this field relies on using L_1 penalised linear regression methods to estimation a sparse precision matrix row by row, the most basic of this being neighbourhood selection. This also allows us to use advances in linear regression techniques to improve precision matrix estimation (for instance, SCIO and the Scaled Lasso). Certain properties can also be exploited to guarantee a positive definite result (e.g. the graphical lasso).

When we compare the methods using the area under the ROC curve, simply thresholding the sample covariance matrix performs very competitively with the sparse precision matrix estimation methods, indicating that there is much work to be done in getting good results out of these methods.

Furthermore, when we look at selecting one particular network using cross validation we do not always achieve a great deal of success. In fact, for the networks with a uniform or power law structure we achieve F_1 scores of less than 0.4 for all methods, indicating how challenging it is not just to have a method that can uncover the underlying truth, but also to select the correct regularization parameter to this end. The threshold method is still competitive with the sparse regularization methods in this situation. However, if we look at the results for the caveman networks, the threshold method performs quite poorly, and the sparse precision matrices well, indicating that the underlying structure of the network matters significantly.

This poor performance when selecting a particular network is likely to be an issue when using these methods for the estimation of financial networks. This application is an example of a situation where we do desire one particular network to analyze, and so it being merely possible to extract the correct network is not sufficient. Furthermore, we do not have a ground truth model of the financial network, meaning that we cannot rely on the good performance of a particular method for a particular network structure.

Chapter 4

Robust Risk Minimization using the Graphical Lasso

4.1 Introduction

Portfolio optimization and its variants have been of interest in empirical finance for decades, following the pioneering work of Markowitz [161]. This is based on the idea of formulating an optimization problem to minimize the risk of a portfolio for a desired return, under the assumption that the log returns of the assets follow a multivariate Gaussian density.

The risk of a portfolio is defined as the variance of the returns. This can be expressed as:

$$\boldsymbol{w}^T \Sigma \boldsymbol{w} \tag{4.1}$$

where $\boldsymbol{w}$ is a vector containing the portfolio weights, and Σ is the covariance matrix of the asset returns. The returns of said portfolio are defined as:

$$\boldsymbol{w}^T \boldsymbol{\mu} \tag{4.2}$$

where $\boldsymbol{\mu}$ is a vector containing the mean return of the assets. Using these we can formulate the optimization procedure:

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \boldsymbol{w}^T \Sigma \, \boldsymbol{w} \\ \text{subject to} \quad & \sum_{i=1}^{i=N} w_i = 1 \\ & w_i \geq 0 \\ & \boldsymbol{w}^T \boldsymbol{\mu} \geq \rho \end{aligned} \tag{4.3}$$

where ρ sets a minimum return. Solving the above quadratic programming problem at different values of ρ yields a set of portfolios with different trade offs between risks and returns.

We desire that the resulting portfolios have several properties. Firstly, the out of sample performance must be good. In this situation, this means a low out of sample risk. Secondly, buying and selling assets is not free. Therefore, we do not want to own large amounts of different assets (referred to as a position), and furthermore when we re-estimate our optimal portfolio, we do not want many changes between the old and new portfolios. The size of the changes required from one window to the next is known as the turnover of the portfolio.

Unfortunately, the estimation of the mean and covariance from financial data is known to suffer from robustness issues [214], [143], [21]. Financial data consists of

occasional outliers to which maximum likelihood estimates are notoriously sensitive. Furthermore, to estimate covariance matrices reliably we need a long enough window (n) of data, though due to non-stationarity in the markets, we may choose a small window. This (particularly when n and p are of similar values) can lead to the covariance matrix Σ being singular and non-invertible. The consequence of poor estimation of parameters is that the resulting portfolio can be unstable and produce poor out-of-sample performance, with extreme weights that are liable to have large changes over time [21]. In many cases these portfolios perform worse than a portfolio where each asset is evenly weighted (referred to as a $\frac{1}{p}$ or naive portfolio) [63], although it has been argued this is due to the use of small windows for parameter estimation [135]. These issues are often addressed by regularization, of which shrinkage estimation is a classic tool (*e.g.* [143]). Brodie *et al.* address this issue by regularization using the l_1 (or *lasso*) penalty in an index tracking setting, deriving stable portfolios which are also sparse. Takeda *et al.* [249] use a combination of l_1 and l_2 regularizers to simultaneously induce sparsity and improve out-of-sample performance. Regularizing the portfolio weights can also be related back to shrinking the covariance matrix in some way [64].

Sparse precision matrix estimators have shown some promise in the financial field, in reducing portfolio turnover when hedging [94], and some work has claimed they can achieve a lower risk when compared to other estimators [6] [257]. In this chapter we evaluate the performance and robustness of portfolios constructed using these estimators

4.2 Methods

We focus on the estimation of the covariance and precision matrices, and so we do not set a minimum desired return. Instead we look at solving the unconstrained minimum variance portfolio, defined as:

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \boldsymbol{w}^T \Sigma \boldsymbol{w} \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{w} = 1 \end{aligned} \tag{4.4}$$

We construct portfolios both with and without a short selling constraint ($w_i > 0$). Short selling refers to selling stocks which the seller does not own. On a practical level, this is not always possible, and as far as the portfolios are concerned, a short selling constraint can actually improve performance [114]. Effectively this constraint acts as a regularizer, as portfolios without this constraint can have very large positive and negative weights. If we do not have a short selling constraint, the problem has an analytical solution:

$$\boldsymbol{w}^* = \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \Sigma^{-1} \mathbf{1} \tag{4.5}$$

If we do have a short selling constraint, the problem does not have an analytical solution.

From equation 4.5, we can therefore see that accurate estimation of the precision matrix is key in selecting a portfolio with good performance. In particular we focus on 3 methods - SCIO [153], Scaled Lasso [246] and the Graphical Lasso [87]. It is important to note that only the Graphical Lasso will give us a precision matrix that is guaranteed to be both symmetric and positive definite. For the other methods we can make them symmetric without much effort, but ensuring they are positive definite while maintaining the sparsity is much more challenging. With a portfolio

that does not have the short selling constraint this is not an issue as we have a formula for the solution (see equation 4.5). Obviously a precision matrix that is not positive definite is not a correct model. This creates a question of whether estimating the true non-zero values or having a more accurate overall model is more important. We also use three non-sparse methods, the sample covariance matrix (which provides a baseline), the Ledoit-Wolf estimator, which is regularised in a different way (see Chapter 3) and portfolios where we invest evenly in every asset (referred to as Naive).

If we do have a short selling constraint then our solver can fail on certain precision matrices. We therefore report when it fails in the results.

4.3 Software and Data

We make use of Python, NumPy [193] and SciPy [118] for general scripting, scikit-learn [201] for the implementation of the glasso and NITK for an implementation of the other precision matrix estimators.

The data we use is downloaded from Yahoo Finance. For the UK data we use the FTSE100 companies, for the US returns we use the S&P500 companies and for the German data we use the DAX30 companies. We use returns from 2000/03/01 to 2019/10/21. For each dataset any company missing more than 10% of its data is removed, and any missing values are filled forwards from the first good value. If the values are missing from the start we backfill from the first good value. We use a window of 252 trading days (1 trading year) for Germany and the UK, and 504 for the US due to the larger sample size. The portfolios are then tested for out of sample performance on the following year (UK and Germany) or two years (for the US)

To select the regularization parameter we use BIC (although note the scaled lasso can calculate its own regularization parameter) since cross-validation cannot be used in an unaltered state with time series data.

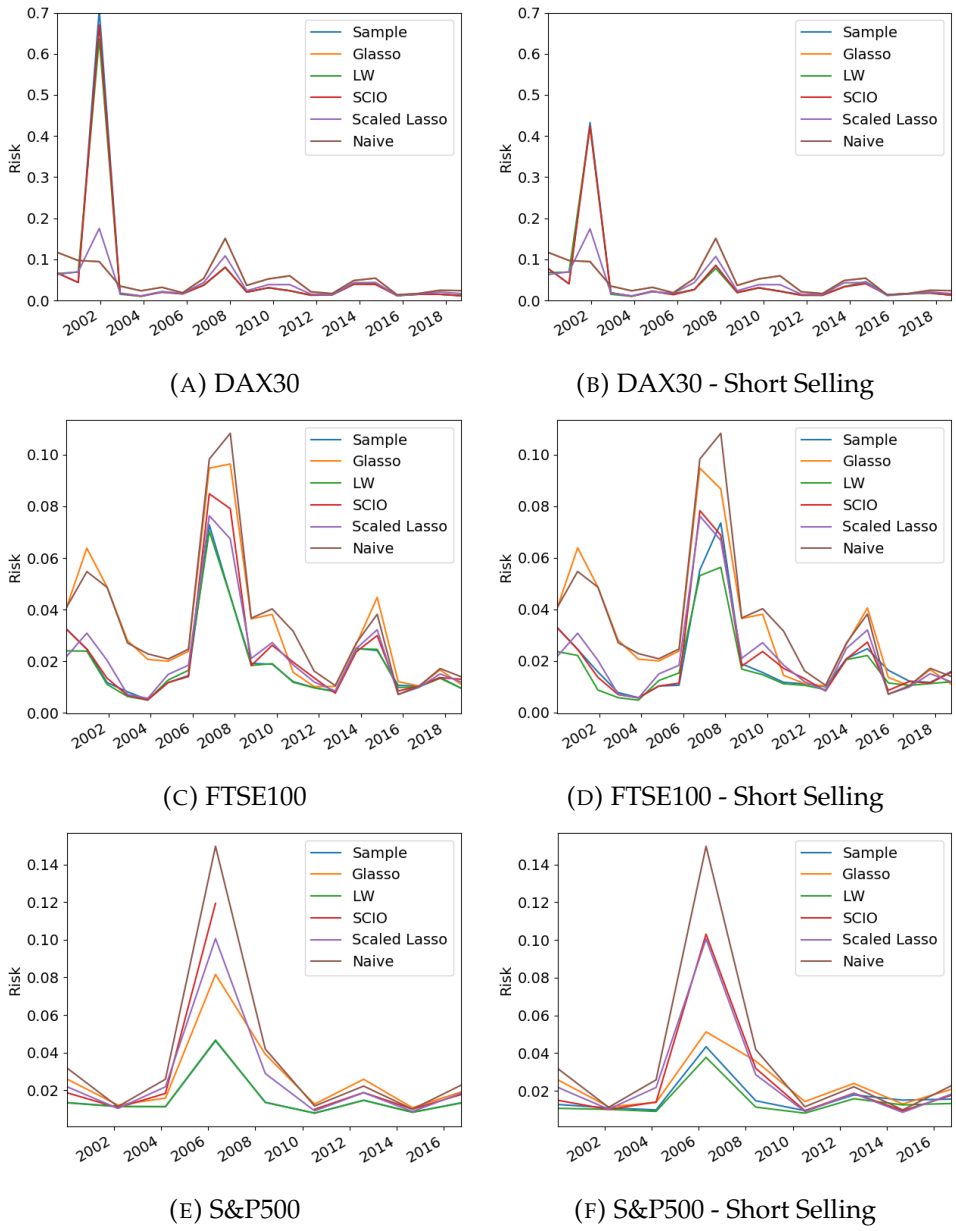
4.4 Results

4.4.1 Portfolio Optimization

To start we compare the out of sample risk of the portfolios. If using daily data we use 252 trading days of data to estimate the covariance matrix and then test the portfolios on the corresponding 252 days of data (252 days = 1 trading year). If using monthly data we use 30 months of training data and 30 months of test data. The sample covariance matrix was not well formed enough at any point to be used without a no short selling constraint so we used the pseudo inverse instead. The risk of the portfolios is annualized. The results are shown in Figures 4.1 (daily) and 4.2 (monthly).

Generally it seems that the Ledoit-Wolf methods have the best performance on the most windows (i.e. the lowest risk). The sparse methods usually have a higher risk than the Ledoit-Wolf or sample covariance matrices, but a lower risk than the Naive portfolios. SCIO is prone to producing covariance matrices that are not positive definite, with the optimization procedure failing at many points for the US data (both daily and monthly)

Buying and selling stocks is not cost free, and holding large positions with leverage is generally regarded as risky and therefore undesirable. Therefore next we look at the size of the active positions of each portfolio and how much buying and selling the



(A) DAX30

(B) DAX30 - Short Selling

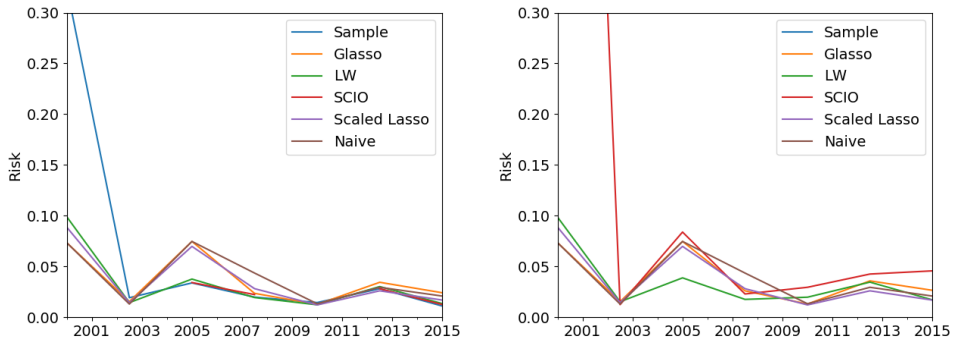
(C) FTSE100

(D) FTSE100 - Short Selling

(E) S&P500

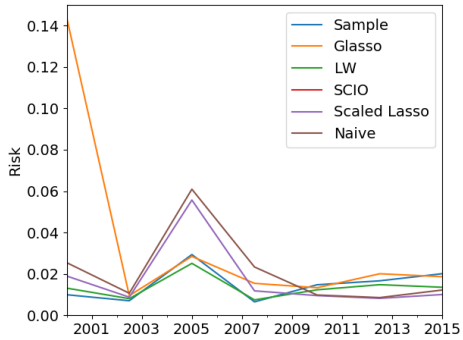
(F) S&P500 - Short Selling

FIGURE 4.1: Test risk of the portfolios on daily data. Figures on the left have short selling forbidden, and those on the right have short selling permitted.

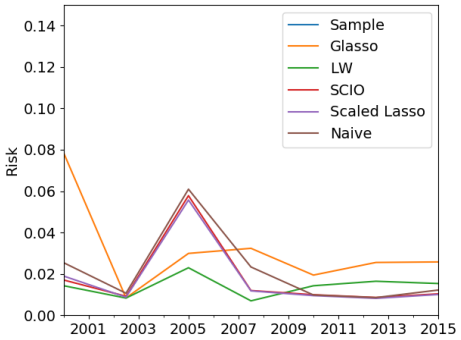


(A) DAX30

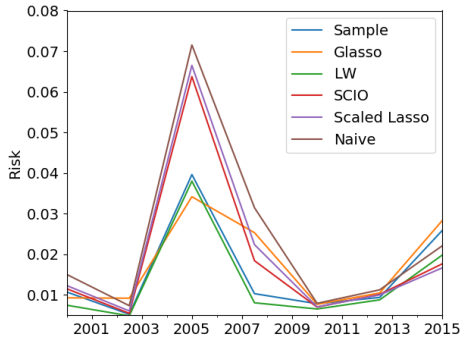
(B) DAX30 - Short Selling



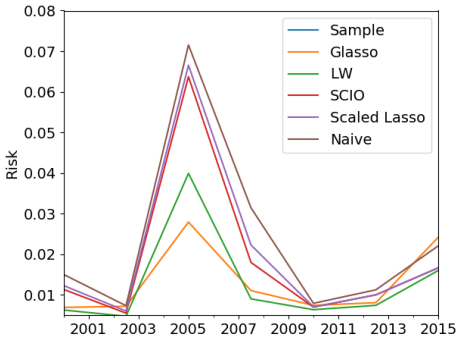
(C) FTSE100



(D) FTSE100 - Short Selling



(E) S&P500



(F) S&P500 - Short Selling

FIGURE 4.2: Test risk of the portfolios on monthly data. Figures on the left have short selling forbidden, and those on the right have short selling permitted.

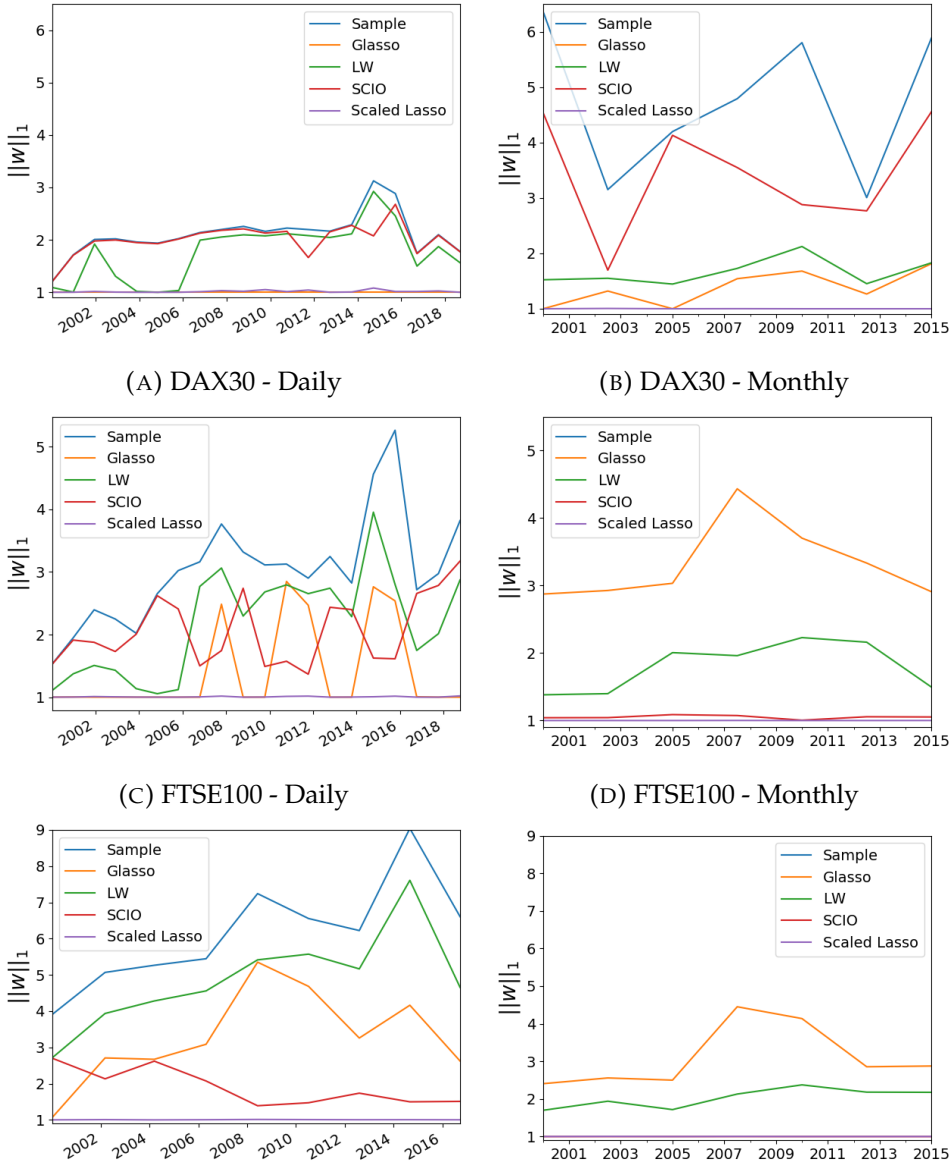
portfolio requires at each re-balancing. We measure the size of the active positions by summing the absolute values of the weights (i.e. the L_1 norm of the portfolio - $\sum_{i=1}^{p} |w_i|$) - a portfolio with no short selling will have an active position size of 1. A smaller position is more desirable. Figure 4.3 shows the active positions of the various methods over the daily and monthly datasets with short selling permitted. For the monthly data for the US and the UK, the sample covariance matrix takes very large positions, so it is omitted from those figures.

Across all of the different countries and time periods, the sample covariance matrix has the largest position. For the larger markets (US and UK), the glasso also tends to take large positions when using monthly data, and the Ledoit-Wolf method also tends to take the second largest positions across most market/time period combinations. Interestingly the scaled lasso tends to have small positions and engage in relatively little short selling. These do show the benefits of regularization in reducing the size of positions a minimum risk portfolio takes. Once we impose the short selling constraint the methods all have an active position of 1, so we omit those figures.

Next we look at the turnover of the portfolios. The portfolio turnover is measured by calculating the difference between adjacent portfolios, and then summing the absolute values of this difference. Again a smaller difference is more desirable - for instance the naive portfolio has a difference of 0 as no buying or selling is ever required (as we have discarded data from companies that are removed from the index in this paper). Figure 4.4 shows the portfolio turnovers over the various datasets for the daily data. The regularized methods provide a significant boost in reducing the portfolio turnover compared to the sample and Ledoit-Wolf covariance estimators when short selling is permitted, with the exception of SCIO. However, when short selling is forbidden the results are less consistent, with only the Scaled Lasso reducing portfolio turnover for all country/time period combinations. Forbidding short selling also reduces portfolio turnover across all estimators. The very low turnover for the glasso portfolios on the German market is due to the parameter selection procedure creating a completely sparse matrix for this market, meaning the portfolios constructed never change.

Next we study this for monthly data. The results are shown in Figure 4.5. Without a short selling constraint the sample covariance matrix was so ill-formed that the positions held were massive and the turnover equally large, making comparisons not particularly possible. Therefore the sample covariance turnover is not presented in the figures for short selling. We see similar results to the daily data, with the sparse estimators generally causing a reduction in the portfolio turnover, but forbidding short selling having the greatest effect. Again the Scaled Lasso has the lowest turnover out of all the methods.

In general our results say that sparse precision matrix estimation methods do not tend to give better results than non sparse methods when looking at the out-of-sample risk (this has also been found when using the sparse precision matrices for classification [217]). Imposing a non-short selling constraint is perhaps the best way of regularizing the covariance matrix, and the Ledoit-Wolf methods were the most consistent in achieving the best results. However the sparse methods can give an improvement with regards to the portfolio turnover and active positions held (if short selling is permitted), with the Scaled Lasso performing well on this measure. It is also well known that lasso based methods (of which all our sparse methods are) do not perform particularly well on highly correlated data [271], which unfortunately financial data is [211] and perhaps this could explain the poor performance of the resulting portfolios.



(A) DAX30 - Daily

(B) DAX30 - Monthly

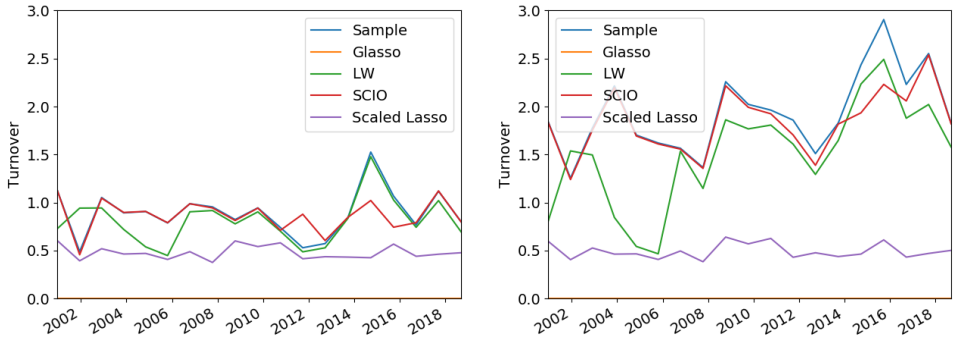
(C) FTSE100 - Daily

(D) FTSE100 - Monthly

(E) S&P500 - Daily

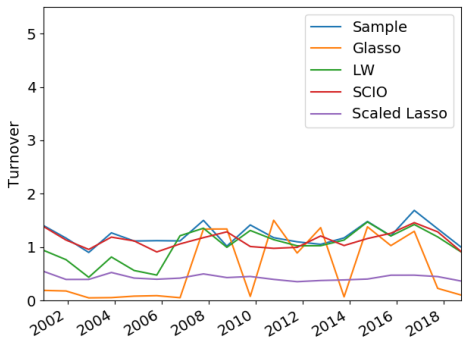
(F) S&P500 - Monthly

FIGURE 4.3: Active positions for the various covariance estimation methods with short selling permitted. Note how the sample covariance matrix has the largest position size of all the methods - regularization helps immensely with reducing this. For the UK and US monthly data, the active position of the sample covariance matrix is very large, so it is omitted.

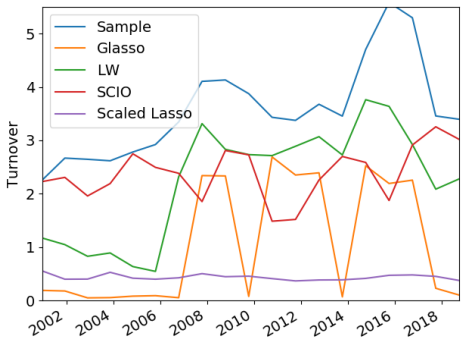


(A) DAX30

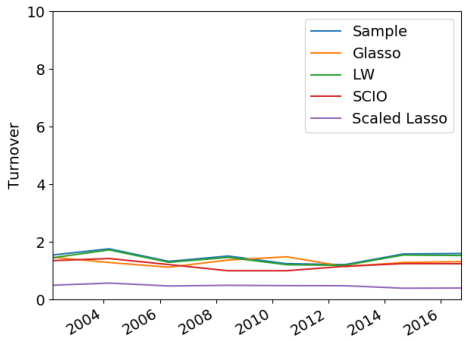
(B) DAX30 - Short Selling



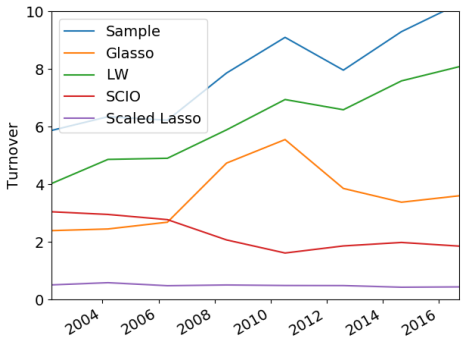
(C) FTSE100



(D) FTSE100 - Short Selling

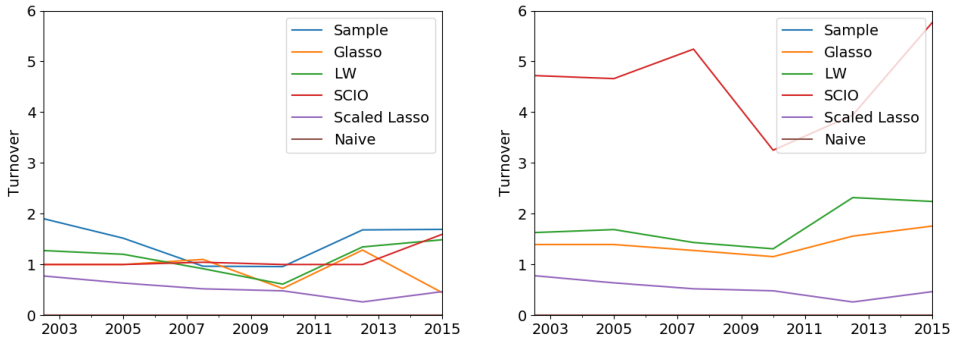


(E) S&P500



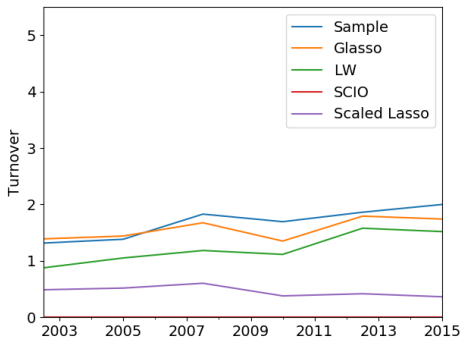
(F) S&P500 - Short Selling

FIGURE 4.4: Portfolio turnover for the various covariance estimation methods on daily data. Regularization helps in reducing portfolio turnover, and the sparse methods do seem to have a smaller turnover than the non-sparse methods, but forbidding short selling has the largest effect.

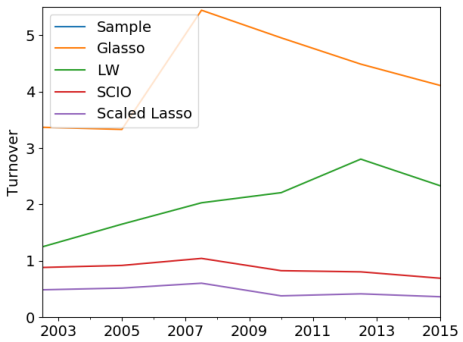


(A) DAX30

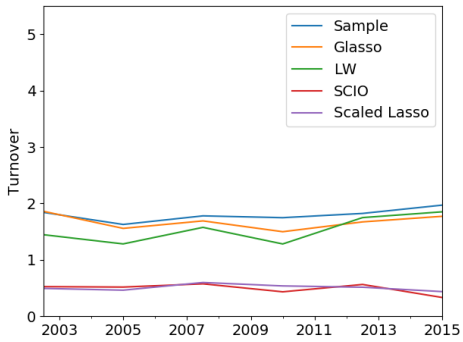
(B) DAX30 - Short Selling



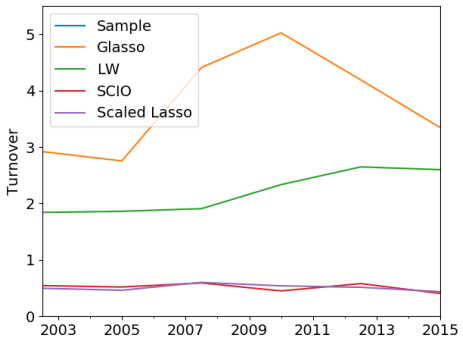
(C) FTSE100



(D) FTSE100 - Short Selling



(E) S&P500



(F) S&P500 - Short Selling

FIGURE 4.5: Portfolio turnover for the various covariance estimation methods using monthly data. The sample covariance is omitted from the short selling results as it is very large. Again banning short selling reduces turnover, and the sparse methods also provide a reduction in portfolio turnover when short selling is permitted.

4.4.2 Portfolio Robustness

In this section we explore the robustness in the performance of the methods - how sensitive they are to slight changes in the input. As mentioned in the introduction, the sample covariance estimator is not robust to the presence of outliers, which occur often in financial returns data. We hope that the regularization in the sparse estimation methods may reduce the effect of outliers, and result in a more robust estimate of the precision matrix. A more robust precision matrix should result in better out of sample performance, and less portfolio turnover between windows.

To measure robustness, we use a circular bootstrap to create multiple sub-datasets from our main one and create minimum risk portfolios using the covariance estimation methods on each sub-dataset. A circular bootstrap is a form of the bootstrap designed for time series data. Rather than randomly picking samples with replacement from the dataset, we select a set of samples continuous in time, and 'wrap around' if we go past the end. This allows us to avoid 'look-ahead' effects in our data. We set our training window as 504 days for the UK and Germany and select 252 days of data from this for each bootstrap. For the US we use 1008 days for a window and select 504 days for each bootstrap. These portfolios are then tested on out of sample data from the following 252 (for UK and Germany) or 504 (for the US) after the test set.

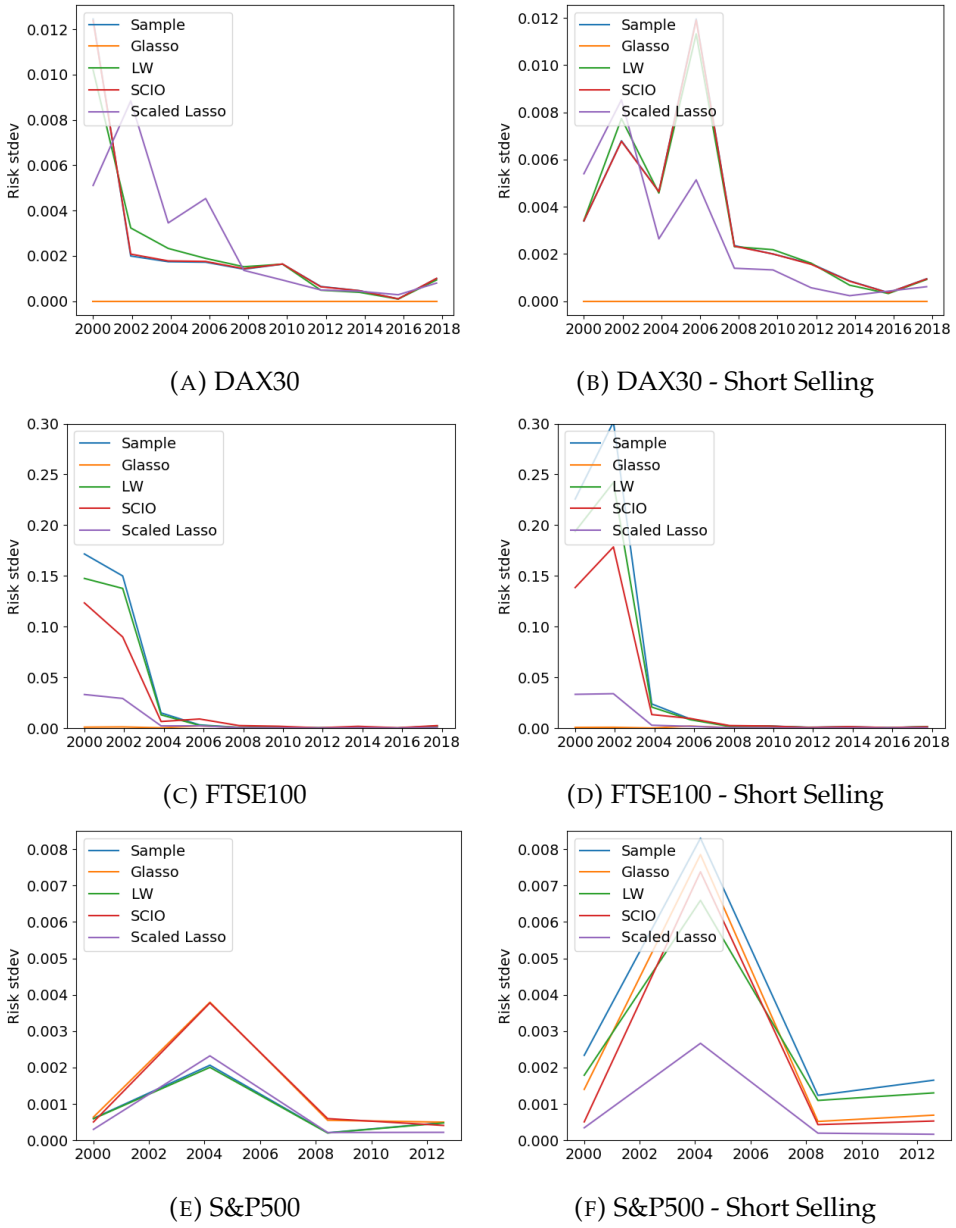
We evaluate the robustness of the covariance estimation methods using the standard deviation of the risks of the bootstrapped portfolios. If a method is robust, the portfolios produced will be similar and therefore the standard deviation of the risk low. The results are shown in Figures 4.6 (daily) and 4.7 (monthly).

From these figures, we firstly note the effect of permitted short selling - in almost all cases this causes an increase in the standard deviation of the risks. In general the Scaled Lasso has the lowest standard deviation across the various time periods and markets. In general we find that the regularized methods can sometimes provide an increase in robustness, but it is method specific and not always the case. We can also see that times of market stress cause a large decrease in the portfolio robustness. For many of the time periods, SCIO failed to produce a positive definite precision matrix, hence why it often has missing values in the graphs.

There are some country specific results. For the German daily data, the glasso selects no non-zero values on the off-diagonal of the precision matrix. This explains the lack of any standard deviation in the risk, but also should not be considered a good result. If short selling is permitted on the German daily data, the results of SCIO and the sample covariance are almost identical.

Also, for the two larger markets, with monthly data, the sample covariance matrix is not positive semi-definite, meaning it cannot be used when short selling is forbidden. This is why it is missing on our figures. Furthermore, when short selling is permitted for the German monthly data, the sample covariance performed very poorly, so it is omitted from the results.

In conclusion, it appears that the Scaled Lasso alone provides an increase in robustness compared to the sample covariance matrix across all of our parameters, countries and time periods. The other estimation methods do sometimes provide an increase in robustness compared to the sample covariance estimator for certain time periods, but this is not consistent and can in fact produce less stable portfolios. Therefore they should not be relied on.



(A) DAX30

(B) DAX30 - Short Selling

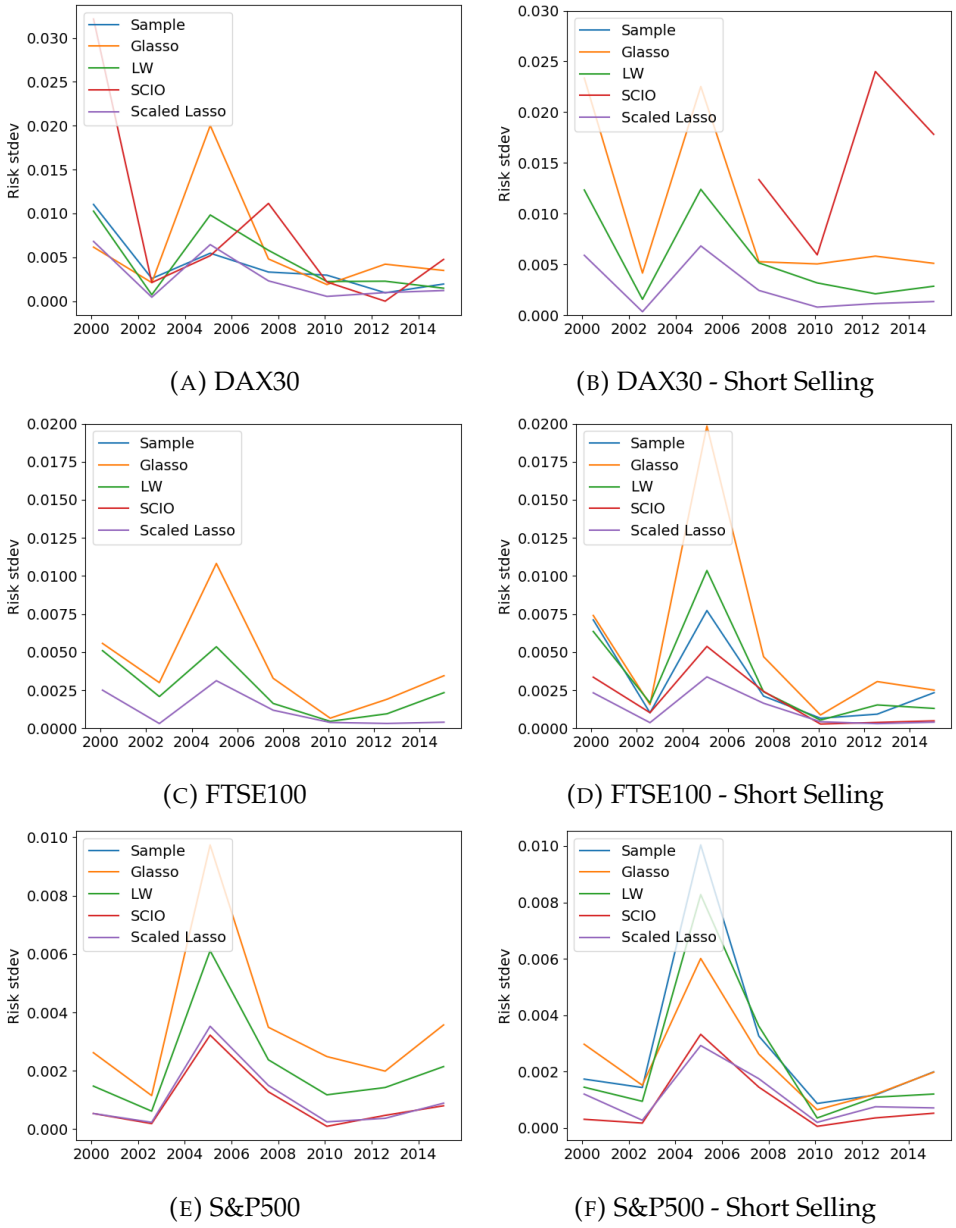
(C) FTSE100

(D) FTSE100 - Short Selling

(E) S&P500

(F) S&P500 - Short Selling

FIGURE 4.6: Standard deviation of the risk of the portfolios created from the daily bootstrapped datasets over time



(A) DAX30

(B) DAX30 - Short Selling

(C) FTSE100

(D) FTSE100 - Short Selling

(E) S&P500

(F) S&P500 - Short Selling

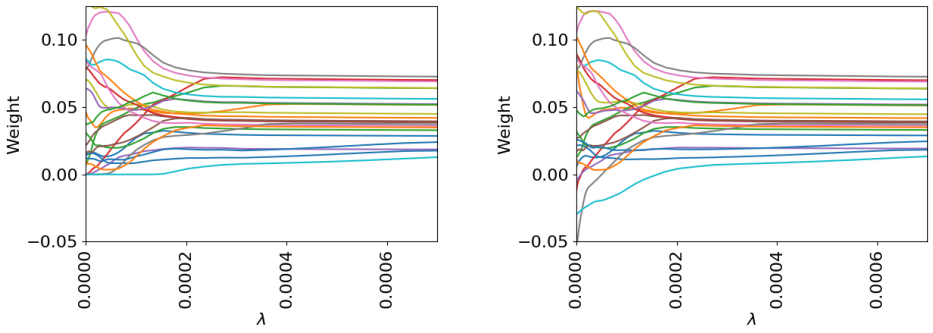
FIGURE 4.7: Standard deviation of the risk of the portfolios created from the monthly bootstrapped datasets over time

4.4.3 Effect of the Regularization Parameter

Throughout this paper we have chosen the regularization parameter in a data driven way, however there is of course the question of how robust the portfolios produced are to changes in this parameter. For simplicity we focus on the graphical lasso here. As we increase the regularization parameter the precision matrix will become more sparse (in our experience if the regularization parameter is increased enough the covariance matrix estimated will also become sparse, but this is not guaranteed). To start we plot the weight on each company in a portfolio for a variety of λs to show how it changes. The maximum value for λ is chosen as the smallest value that will give us a completely sparse precision matrix. This can be calculated analytically, it is the largest absolute value in the off-diagonal of the sample covariance matrix [167].

$$\lambda_{\max} = \max_{i \neq j} |S_{ij}| \tag{4.6}$$

We then choose $\lambda_{\min} = 10^{-3}\lambda_{\max}$ as recommended by Friedman et al. [88], and use log spacing from $\lambda_{\min}$ to $\lambda_{\max}$. The DAX30 is used as it has a small number of companies, making this effect easier to see. The results are shown in Figure 4.8 (the top shows short selling permitted and bottom short selling prohibited)



(A) Short selling forbidden (B) Short selling permitted

FIGURE 4.8: Effects of increasing λ on portfolio weights constructed from DAX30 data with short selling forbidden (left) and permitted (right). As we increase λ the portfolios become more diverse, with more companies being invested in. We can see the portfolio weights moving together as λ is increased.

From this we can see the portfolios become more diverse as we increase λ and spread their weights out more evenly. Furthermore short selling stops with high values of λ even if it is permitted in the optimization procedure. Eventually as we reach a precision matrix where the entire off-diagonal is 0, the investments will simply be inversely proportional to the variance of the stock. If we wish to force the weights to be completely equal, we would have to start adding λ to the diagonal of the sample covariance matrix.

If we look at the solution to the unconstrained optimization problem, we can see why this would be the case. The solution is:

$$\boldsymbol{w}^* = \frac{1}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}\Sigma^{-1}\mathbf{1} \tag{4.7}$$

If we ignore the normalization part ($\frac{1}{\mathbf{1}^T\Sigma^{-1}\mathbf{1}}$) we get that the weight assigned to a stock is the sum of its row in the precision matrix

$$\boldsymbol{w}^* \propto \Sigma^{-1}\mathbf{1} \tag{4.8}$$

$$w_i^* \propto \sum_{j=1}^{p} \Sigma_{ij}^{-1} \tag{4.9}$$

Therefore if we penalize the norm of each row (like these sparse estimators do, although note they usually do not penalize the diagonal) we will cause the weight invested in a stock to decrease. Furthermore if the norm of a row is large, it will cost more than if the norm of a row is small, so large investments are more expensive than small ones. Since it is renormalized to ensure the total weights add to 1, effectively we reduce large investments more than small ones.

Next we measure how the train and test risks and returns vary as we increase λ, experimenting with both short selling permitted and disallowed. The difference in train and test risks are shown in Figure 4.9 (short selling permitted) and 4.10 (short selling forbidden). Both train and test risks appear to increase as we increase λ, which is perhaps to be expected as we are giving the model less flexibility to select a low risk model. However there is a plateau for the test risks, where actually very small values of λ generate a slightly higher risk model than for slightly larger values. This could be due to the procedure overfitting on the precision matrices with low regularization. We can also see that allowing short selling allows the train risk to drop very low, but this is not reflected in the test risk, which is roughly similar irrelevant of whether short selling is permitted.

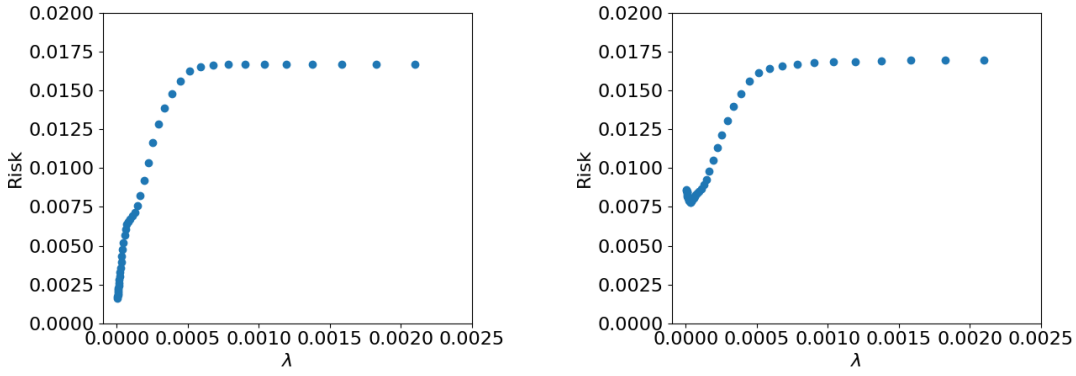


FIGURE 4.9: Train (a) and Test (b) risks of the portfolios constructed from S&P500 data as we vary λ with short selling permitted.

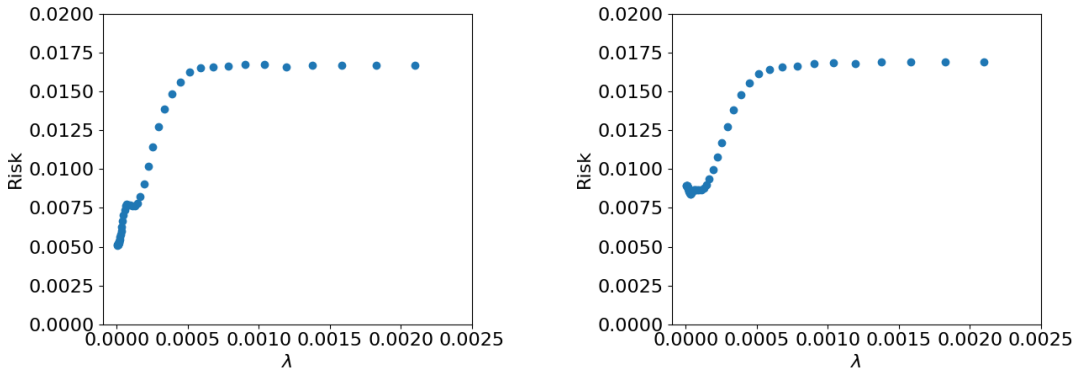


FIGURE 4.10: Train (a) and Test (b) risks of the portfolios constructed from S&P500 data as we vary λ with short selling forbidden.

4.5 Conclusion

In this chapter we have compared precision matrices estimated using sparse methods (i.e. the graphical lasso and its variants) to the sample covariance matrix and the Ledoit-Wolf shrinkage estimator by using them to construct low risk portfolios. These portfolios are constructed from daily and monthly returns from the German, US and UK markets and we experiment with both forbidding and allowing short selling. To evaluate the portfolios we compare their risk on out of sample data. In general we do find that the Ledoit-Wolf shrinkage based method has the best performance on the out of sample data, and that simply enabling short selling generally means the sample covariance matrix also out performs the sparse methods. However the sparse methods can provide an improvement when it comes to portfolio turnover - the portfolios produced are more stable.

Next we investigate how robust the various estimators are by using a circular bootstrap on the datasets and looking at the mean and standard deviation of the out of sample portfolios. In general we do not find any one estimator consistently out performing the others on all of the different combinations of country and time period. We do however note that adding a short selling constraint also improves the robustness of the sample estimator.

Finally we investigate how the regularization parameter for the graphical lasso effects the weights and the train and test performance of the portfolios. Increasing the regularization parameter decreases the number of off-diagonal non-zero values until eventually there is only the diagonal left. At this point this optimization procedure will merely invest proportionally to the inverse of the variance for each stock. Interestingly increasing the regularization parameter reduces the amount of short selling and causes the portfolio investment to be more evenly distributed among the stocks. It also has relatively little effect on the actual risks and returns of the portfolios. We would contrast this to the work of Brodie et al. [44], who focus on the estimation of sparse portfolios. In our case, a sparse precision matrix will lead to a dense portfolio.

From our experiments, we would not recommend the use of these sparse estimators for portfolio optimization, as we have not found that they can be relied on to consistently out-perform the Ledoit-Wolf estimator.

Future work on this could involve allowing the penalization of the diagonal as well as the off-diagonal. While this might make the matrix less well formed, it could allow sparsity within the portfolio weights by forcing a row to 0. Alternatively elastic net based [294] methods could be used. These perform much better on correlated data. However this does then give us two parameters to estimate rather than one, which generally tends to be challenging to achieve.

Chapter 5

Partial Correlation Financial Networks

5.1 Introduction

In this chapter we present an empirical study on the inference and analysis of partial correlation networks using a dense partial correlation estimator.

We hope that by using partial correlation we can discover latent relationships that are hidden in correlation networks by the overall movement of the market (the market mode). Previous work has focused on removing the market mode from the data by various methods, including deleting the largest eigenvalue and eigenvector, or using a factor model [182] but to our knowledge a partial correlation approach like the one we take has not been used on stock return data before. In our approach, the partial correlation between two variables is the correlation between the two once the linear effects of all other variables have been removed. To gain further insight, first consider that we can infer row i of a precision matrix from a dataset X via regressing variable i ($\boldsymbol{x_i}$) on the others (X_{-i})

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} ||\boldsymbol{x_i} - X_{-i}\boldsymbol{\beta}||_2^2 \tag{5.1}$$

The solution to this is

$$\beta_{ij}^* = -\frac{\Theta_{ij}}{\Theta_{ii}} = P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}} \tag{5.2}$$

This means that the partial correlation between i and j (P_{ij}) is proportional to the weight that the least squares method would assign to j in a regression problem if we tried to predict i from the rest of the dataset (or since $P_{ij} = P_{ji}$ the weight that the least squares method would assign to i if we tried to predict j from the dataset)

Furthermore the precision matrix is intimately connected with the minimum variance portfolio. The problem, formulated as [161]

$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \boldsymbol{w}^T\Sigma\boldsymbol{w} \\ \text{subject to} \quad & \mathbf{1}^T\boldsymbol{w} = 1 \end{aligned} \tag{5.3}$$

has a solution

$$\boldsymbol{w}^* = \frac{1}{\mathbf{1}^T\Theta\mathbf{1}}\Theta\mathbf{1} \tag{5.4}$$

where $\mathbf{1}$ is a vector of all 1s and w_i is the amount to be invested in asset i. Due to the relationship between the precision matrix (Θ) and partial correlation (P) (see equation 5.2) we expect these networks will give insight into the benefits and drawbacks

of these portfolios, however it is important to remember a partial correlation matrix discards the diagonal of the precision matrix

As previously mentioned, we require the covariance matrix to be invertible to obtain the precision matrix and the partial correlation matrix. This is not always the case with financial data. To solve these issues we use the Ledoit-Wolf Covariance shrinkage method [142]. This produces a well-regularized covariance matrix by combining the sample covariance matrix with the identity to reduce the off-diagonal values. This well formed covariance matrix can then be inverted to obtain the precision matrix. With these regularized covariance and precision matrices, we can use the equations above to obtain the correlation and partial correlation networks. Once obtained, we can then compare and contrast said networks. The shrinkage method is described further in section 5.2.

Our approach in this chapter is to compare these partial correlation networks to standard Pearson correlation networks, and to see if the partial correlation networks can extract structure that is not present in said correlation networks. We also hope that there will be some stable structure, usually covered up by the market mode, that will be present for longer periods of time. While this is a subjective measure, previous work has shown the presence of long term correlations and structure in the financial markets [172] [262] [264], and with the lack of ground truth models, attempting to detect this is the most common approach. If we cannot detect any long term structure, this indicates that most of the edges in the networks are noise.

5.2 Ledoit-Wolf Covariance

Ledoit-Wolf covariance is based upon shrinkage, where we combine the sample covariance matrix S (which may have a high variance but low bias) with a known matrix with desirable properties (which has a low variance but high bias). The usual choice for this is the identity matrix I and we create a linear combination of the two

$$\Sigma_{\text{lw}} = (1-\rho)S + \rho \text{tr}(S)I \tag{5.5}$$

To decide ρ we wish to minimize the Frobenius norm of the difference between Σ_{lw} and the true population covariance matrix Σ_*

$$\min_{\rho} E[||\Sigma_* - \Sigma_{\text{lw}}||_F^2] \tag{5.6}$$

The optimal solution of ρ is

$$\rho = \frac{E[||S - \Sigma_*||_F^2]}{E[||S - \text{tr}(S)I||_F^2]} = \frac{\beta^2}{\delta^2} \tag{5.7}$$

The interpretation here is that if S is very close to Σ_* (i.e. our estimate of the covariance is good) then we do not need to shrink much, or if our shrinkage choice does not seem accurate then we should not shrink much either. However the obvious flaw so far is that we need to know the true population covariance matrix to obtain the correct value for ρ - and if we did then we would not need to bother estimating it to begin with! We therefore require estimates of β^2 and δ^2. We can estimate δ^2 as following:

$$\hat{\delta^2} = ||S - \text{tr}(S)I||_F^2 \tag{5.8}$$

and β^2 as

$$\hat{\gamma^2} = \frac{1}{n^2} \sum_{k=1}^{n} ||x_k x_k^T - S||_F^2 \tag{5.9}$$

$$\hat{\beta^2} = \min(\hat{\delta^2}, \hat{\gamma^2}) \tag{5.10}$$

$$\hat{\rho} = \frac{\hat{\beta^2}}{\hat{\delta^2}} \tag{5.11}$$

The constraint on $\hat{\beta^2}$ ensures that $\rho < 1$. While it is rarely necessary, it does help stop us accidentally making our estimate less well formed.

The Ledoit-Wolf method is guaranteed to give us a positive-definite invertible matrix, which is critical in this application as we require the inverse of the covariance matrix (the precision matrix) to acquire the partial correlation matrix.

This covariance estimation method has been applied to the genomics field [224], for portfolio optimization [142] and in the neuroscience field [42] but to our knowledge has not actually been applied to create financial networks.

5.3 Data and Software

For our study we use daily log returns from the S&P500. If there is less than 10% of the data missing for a particular stock we fill it with the price from the previous day, or if the data is missing from the start, from the first day when the stock is traded. If there is more than 10% missing we discard the data for that stock. We use the close price on the day to calculate the return, from 2000-01-03 to 2017-12-05. Overall we have 4510 days of return data for 345 stocks. Since financial data is non stationary we use a window of 300 days and slide along this 30 days at a time to obtain a sample where we can assume the data is stationary, giving us 140 windows overall. The returns inside each window are normalized using the z-score to have a mean of 0 and a standard deviation of 1. While correlation is by definition normalized, this procedure is mostly for the benefit of the shrinkage procedure - normalizing reduces the amount of shrinkage required which allows us to capture more relationships.

Using this dataset we infer a network for each window by using the Ledoit-Wolf shrinkage methods to obtain a covariance matrix and inverting it to obtain a precision matrix. We then scale both of these matrices appropriately to create the correlation and partial correlation matrices and use these as adjacency matrices to construct the networks. We then study the properties of these networks and how they change over time.

We make use of Python, NumPy and SciPy [193] for general scripting, pandas [169] for handing the data, sklearn [200] for the implementation of the Ledoit-Wolf estimation methods, statsmodels [227] for some of the statistical analysis, matplotlib [110] for plotting, Networkx [96] for the network analysis and gephi [16] for some graph visualization.

If the reader is interested in reproducing our work, the code and data can be found at `https://github.com/shazzzm/partial_correlation_financial_networks/`.

5.4 Results and Analysis

5.4.1 Network Analysis

Firstly we display the networks constructed on the first window of this data. Since they are dense, we display the edges that correspond to the 1000 largest absolute values from the off-diagonal of the matrix. The correlation network has isolated nodes in this situation, so we only display the largest connected component, but the partial correlation network remains connected. These networks are displayed in Figures 5.1 (Correlation) and 5.2 (Partial Correlation). Both networks show a degree of sector clustering but it is far more prominent in the correlation networks compared to the partial correlation networks. The partial correlation networks also seem to have a more uniform degree distribution than the correlation ones, with less community structure.

To begin our analysis we look at the distribution of correlation coefficients to partial correlation coefficients in the network, and the difference in weight between the same edge in the two networks. A histogram of these is shown in Figure 5.3, and a scatter plot relating the two is shown in Figure 5.4. In general, partial correlation coefficients tend to be smaller than the corresponding correlation values and are more likely to be negative, but it is also clear that the two are related. This is likely to be due to the definition of partial correlation - if it is reducing the value of indirect correlations then we would expect some companies that are supposedly correlated to have these relationship strengths reduced.

Our next goal is to compare and contrast stability of the networks. In a correlation matrix the largest eigenvalue measures the intensity of the correlation present, and the corresponding eigenvector measures the 'market mode' and the effect the general market has on that particular company [206] [182]. Each entry of this eigenvector can also be used as a measure of centrality. Therefore we can study how this eigenvector changes over time to see if the networks regard the same nodes as important, a proxy for how stable the networks are. To measure this we normalize the eigenvectors so the components to add to 1 and then measure the difference between those from adjacent windows using the L_2 norm.

Firstly we look at how the largest eigenvalue varies over time. The results are shown in Figure 5.5. From this we can see that the largest eigenvalue of the partial correlation matrix is much smaller and varies relatively little, compared to the largest eigenvalue of the correlation matrix. This implies that the intensity of the partial correlation networks does not change much over the dataset, particularly compared to the correlation networks which have large changes. This perhaps indicates that the market mode has been removed. However if we look at the difference in the eigenvectors we get a slightly different story. From Figure 5.6 we can see there is a larger change in the corresponding leading eigenvector of the partial correlation matrix as opposed to the correlation matrix, signifying the partial correlation networks are less stable than the corresponding correlation networks and could indicate as to why minimum risk portfolios tend to require large changes in asset holdings [63]. Both seem to reflect macroeconomic changes, with the magnitude of the difference varying over time. Interestingly the differences between eigenvectors from adjacent windows drops during periods of disruption.

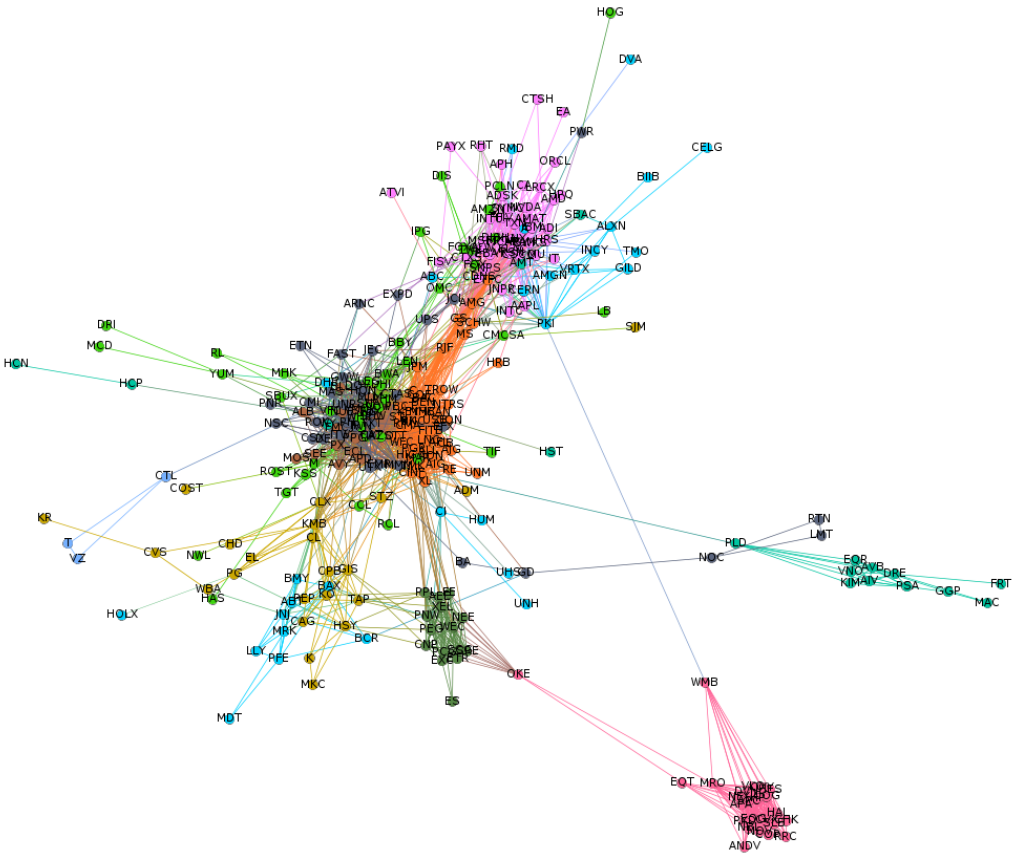


FIGURE 5.1: Example correlation network inferred from the first window. Only the largest connected component of the network containing the 1000 edges with the largest absolute weights are shown. Nodes are coloured according to sector membership. There is a strong community structure visible, with communities usually made up of companies in the same sector.

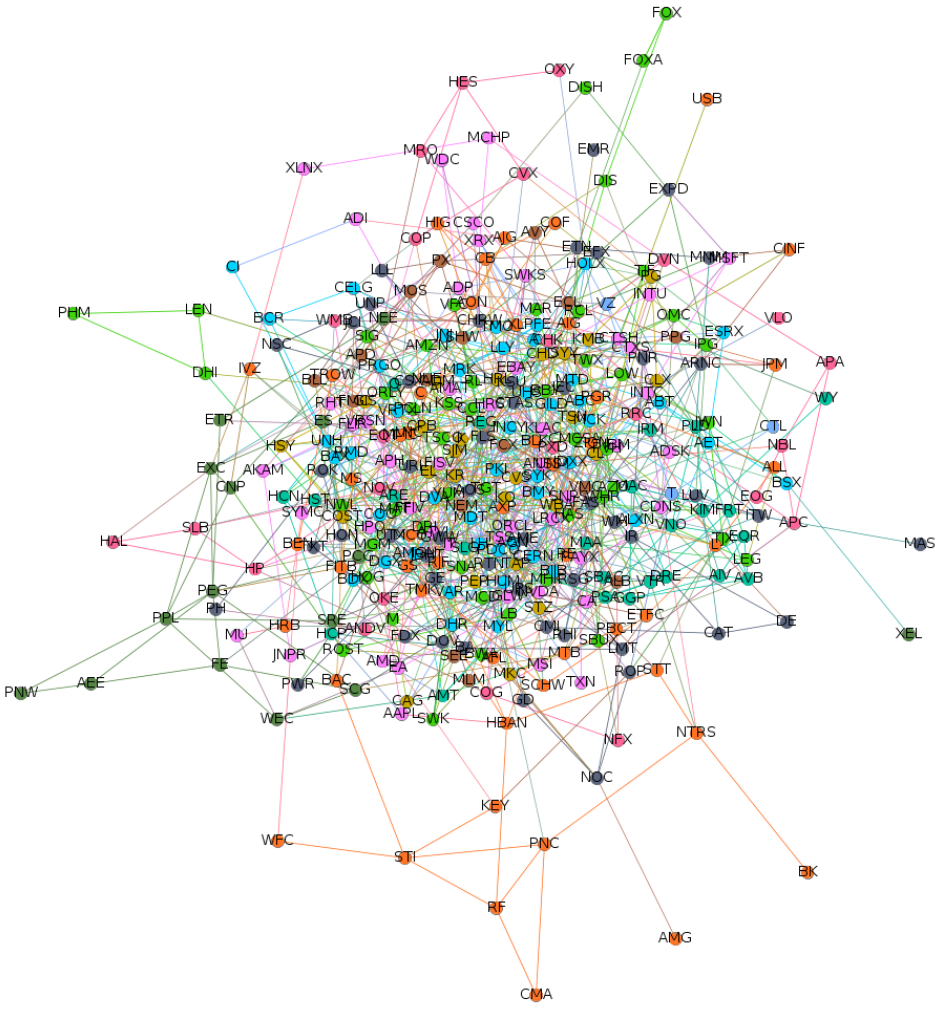
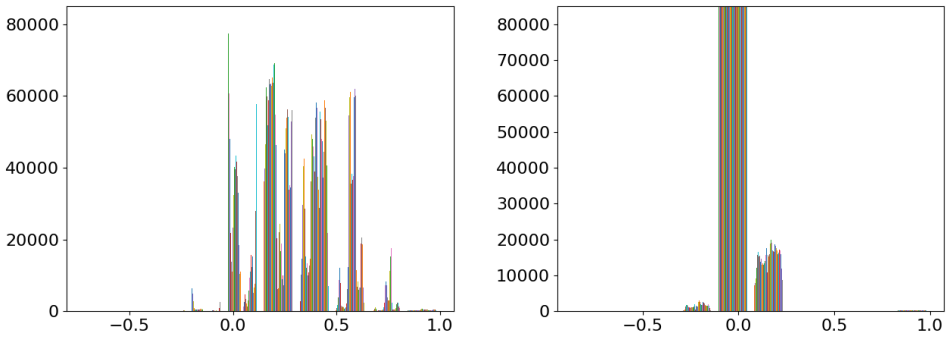


FIGURE 5.2: Example partial correlation network inferred from the first window. Only the 1000 edges with the largest absolute weights are shown. Nodes are coloured according to sector membership. There seems to be less sector clustering in the partial correlation networks and less of a community structure.



(A) Correlation (B) Correlation

FIGURE 5.3: Distribution of correlation and partial correlation coefficients over the dataset of 140 windows taken over the 17 year period. We can see that the partial correlation matrix generally has smaller values than the correlation matrix, and that they are more likely to be negative.

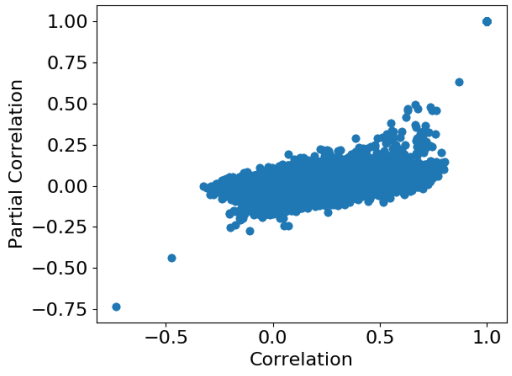
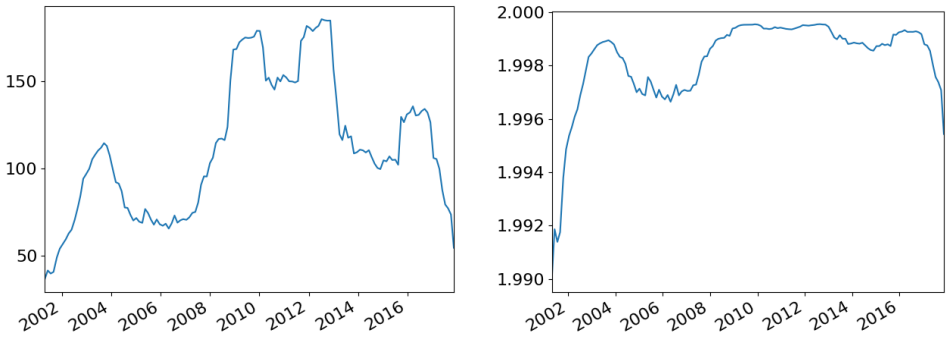


FIGURE 5.4: Scatter plot of the correlation coefficient for an edge against that of the partial correlation coefficient for each of the 140 networks in the dataset. The partial correlation coefficients are in general smaller than their corresponding correlation coefficients which is to be expected if the indirect correlations and reduced, however there is still a relationship between the two



(A) Correlation (B) Partial Correlation

FIGURE 5.5: Largest eigenvalue in the correlation (a) and partial correlation (b) networks. There is a large variation in the largest eigenvalue of the correlation matrix, with it varying from 180 to 40. It noticeably picks up the financial crisis of 2008/2009, where the eigenvalue reaches its maximum. The reverse is true in the partial correlation networks where the largest eigenvalue stays roughly constant (and quite small) showing the network has a consistent intensity, indicating the market state has been removed

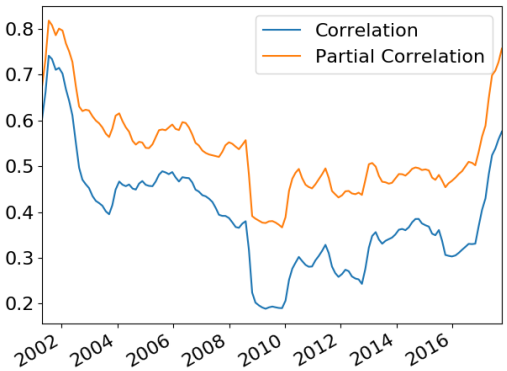


FIGURE 5.6: Change in the normalized leading eigenvectors, measured using the L_2 norm, of the correlation and partial correlation matrices. Both have changes that seem to reflect the general market conditions, although the eigenvector from the partial correlation matrix seems to change more, which shows the network is less stable.

5.4.2 Sector Centrality

Next we study the centrality of the sectors in the networks, allowing us to quantify their influence in the economy. We use two measures, degree centrality and eigenvector centrality. Relating these specifically to financial networks, firstly we note that degree centrality is simply the sum of the edges of a node. The weight on the optimal portfolio is also proportional to weight of the edges on a node (see equation 5.4). Secondly eigenvector centrality is calculated using the eigenvector that corresponds to the largest eigenvalue, with its components normalized to sum to 1 in the same manner as above. This largest eigenvector reflects the market mode and the effect the general market has on a particular company.

The presence of negative edges in the networks makes calculating centrality more challenging. Negative edges can result in a node having negative centrality, which does not have an obvious interpretation. We can use the absolute values of the edges to solve this problem but this involves discarding the negative relationships, which are numerous in the partial correlation networks. In our experiments we found relatively little difference between permitting negative edges or using the absolute values of edge weights and so permit the existence of negative edges. We normalize at the end so the sum of all node centralities is 1.

To measure the centrality of a sector we take the mean centrality of all the companies in said sector. We then normalize these mean sector centralities to add to 1 to make comparison easier. To start with we look at the mean degree centrality for the sectors for each network. A graph of this over time is shown in Figure 5.8. In this graph we can see that the partial correlation networks have a much lower difference and variance in the sector centrality than the correlation networks - each sector has roughly the same centrality and there is little variation over time. The telecommunications sector has relatively few companies, hence why its centrality has far more variance. In the correlation networks we firstly see a much larger variance in the mean centrality of a sector, with the financial sector having the highest mean centrality for the majority of the dataset. Interestingly all the centralities 'jump' together during the financial crisis, showing how suddenly the correlations between previously unrelated companies increases due to these macroeconomic effects.

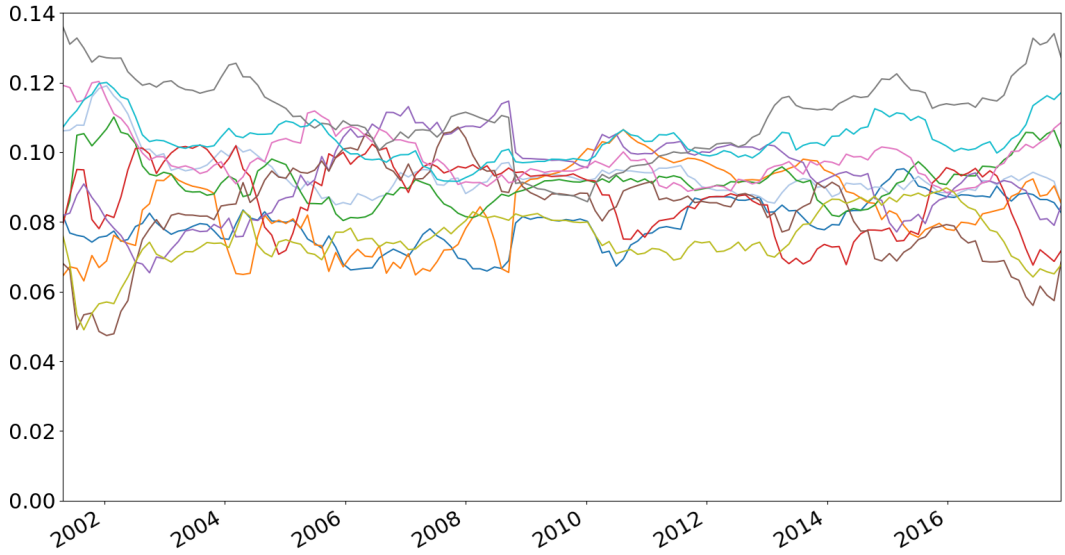
FIGURE 5.7: Legend for the sector colours for Figures 5.8 and 5.9

Next we look at eigenvector centrality. A graph of this over time is shown in Figure 5.9. In this figure we see very different results to degree centrality for the partial correlation networks, which show a much larger variation in the centrality of the sectors. They also show a slightly larger variance in the mean centrality when

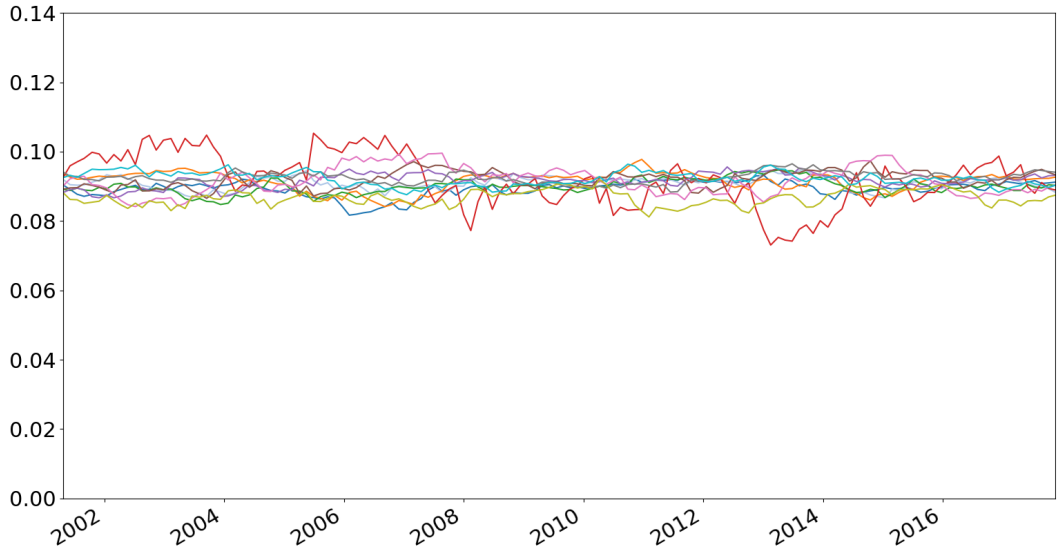
compared with the correlation networks too. In particular the financial sector is far more dominant than we would expect. The macroeconomic conditions are also visible in these graphs, with the financial crisis again forcing the mean centralities towards a mean.

The movement of centrality measures towards a mean during times of disruption is particularly interesting. Preis et al [211] pointed out that the market tends to be more correlation during times of market disruption, which makes selecting truly diversified portfolios very challenging as suddenly supposedly unrelated assets become related during these periods. This may be relevant here too - here we have that during periods of disruption companies start having far more similar behaviour than they did during times of stability.

We have pointed out this connection between the minimum risk portfolios and the degree centrality in the partial correlation networks, but to what degree does it hold? To further explore this we plot the L_2 difference between the optimal portfolio vector and the sum of the diagonal of the precision matrix (as this is effectively discarded by the partial correlation matrix) in Figure 5.10. From this we can see the difference can be quite large, although most of the difference seems to be explained by the size of the precision matrix diagonal, which is effectively discarded by the partial correlation networks.

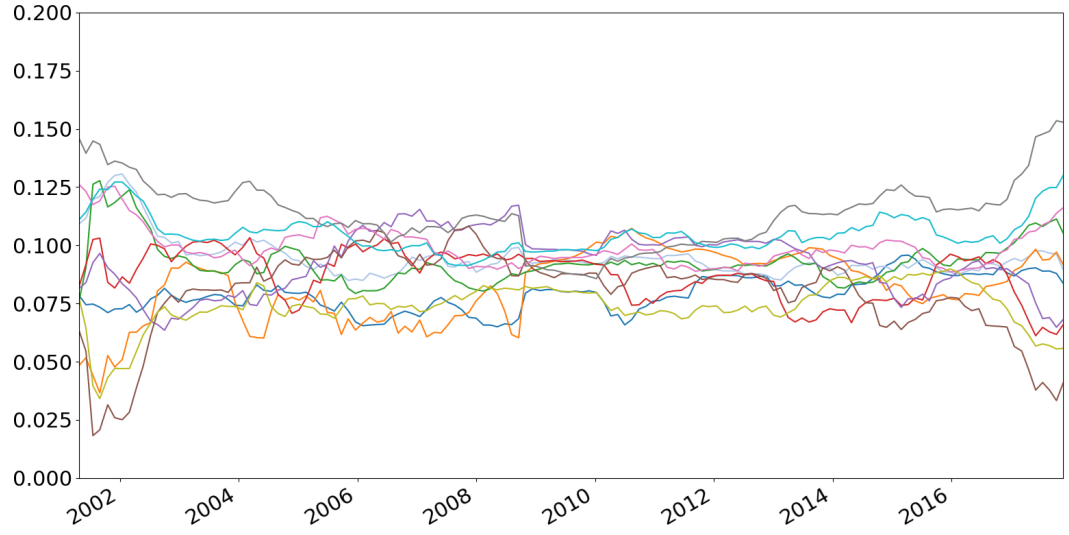


(A) Correlation

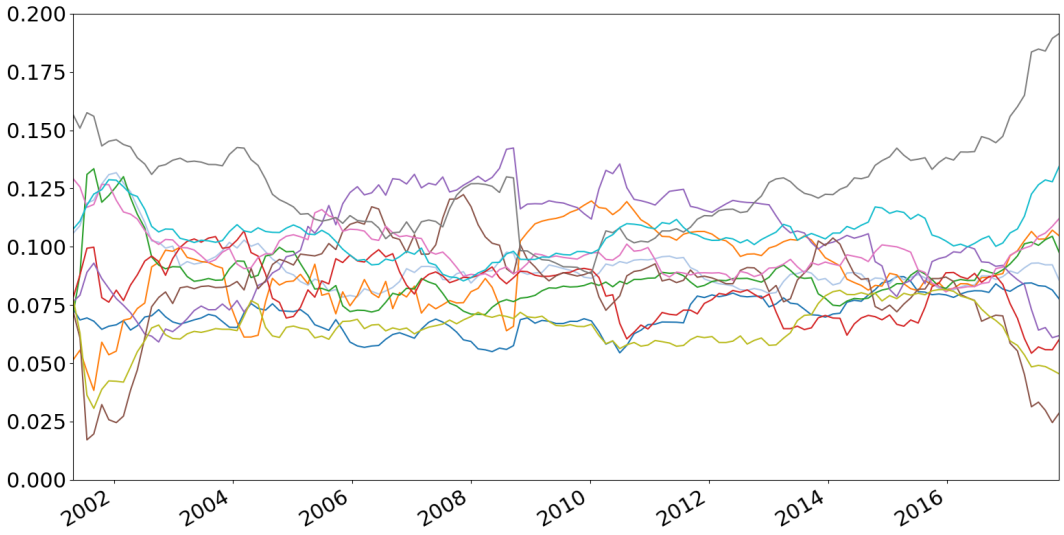


(B) Partial Correlation

FIGURE 5.8: Mean degree centrality for each sector over time. It is noticeable that in the partial correlation networks the difference in centrality is much smaller for each sector than in the correlation networks. We can see the macroeconomic trends in the correlation networks with the centralities jumping together during the crash. The colour legend can be found in Figure 5.7

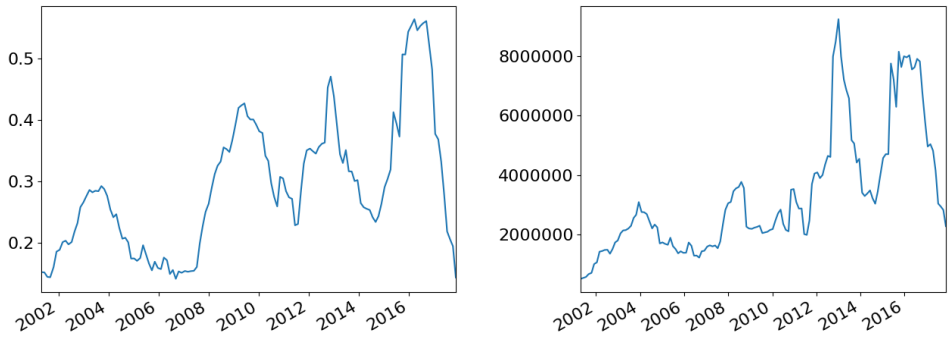


(A) Correlation



(B) Partial Correlation

FIGURE 5.9: Mean eigenvector centrality for each sector over time. These centralities have a much larger variance than the degree centralities, particularly for the partial correlation networks. Here we can see the financial sector is the most central for the majority of the dataset, although the real estate does also become important in both networks. Macroeconomic effects are also much more visible, with the strong change from 2009 - 2011 as all the sector centralities move together. The colour legend can be found in Figure 5.7



(A) Optimal Portfolio weight vs Degree Centrality of node

(B) Sum of the precision matrix diagonal

FIGURE 5.10: L_2 difference between the weight placed on each node in the optimal portfolio vs degree centrality of each node (left) and the sum of the precision matrix diagonal (right). The differences between the optimal portfolio and the degree centrality can be quite large, but most of the difference can be explained by the sum of the precision matrix diagonal.

Sharpe Ratio	Degree	Eigenvector
Correlation	0.156	0.148
Partial Correlation	0.129	0.131
Risk		
Correlation	-0.189	-0.175
Partial Correlation	-0.145	-0.156

TABLE 5.1: Spearman correlation between the centrality measures and the out of sample risk and Sharpe ratio. In all situations there is a positive relationship between centrality and Sharpe ratio, and a negative relationship between centrality and out of sample risk. All correlations are statistically significant at $p < 0.05$ but the correlation is relatively mild.

Sharpe Ratio	Degree	Eigenvector
Correlation	0.154 ± 0.058	0.151 ± 0.060
Partial Correlation	0.127 ± 0.057	0.131 ± 0.057
Risk		
Correlation	-0.189 ± 0.059	-0.175 ± 0.060
Partial Correlation	-0.145 ± 0.060	-0.150 ± 0.061

TABLE 5.2: Results of using a bootstrap to test if there is a significant difference between the networks for the correlation between out of sample Sharpe ratio and centrlaity. The correlation networks have a slightly stronger relationship between centrality and out of sample Sharpe ratio, and a more negative relationship between centrality and out of sample risk.

5.4.3 Out of sample Portfolio Performance

Previous work with correlation networks has stated that companies who are on the fringes of the network have a better Sharpe ratio than those who are more central [204] [210]. We are curious as to how this applies in these networks. Therefore we study the centrality of a company against its out of sample Sharpe ratio (defined as the mean return over the standard deviation of the returns $\frac{\mu}{\sigma}$) and its risk (defined as the standard deviation of the returns) for the next window.

Using Spearman correlation we find that there is mild positive correlation between the centrality of a company and its out of sample Sharpe ratio in every network, and perhaps unsurprisingly mild negative correlation between out of sample risk and centrality. The exact results are shown in Table 5.1. All results are statistically significant at $p < 0.05$.

We verify if there is any significant difference between the two using a bootstrap. We take 300 samples at a time from the dataset and run this for 1000 times. We then calculate the mean and standard deviation of the correlations from the dataset. The results for this are shown in Table 5.2. In this we can see the correlation networks have a stronger relationship between centrality and out of sample Sharpe ratio, and a more negative relationship between centrality and out of sample risk. The standard deviation for both is quite similar.

5.4.4 Community Detection

In Figures 5.1 and 5.2 it can be seen that there is some community structure in the networks, so we are interested in further studying this. To do so we use a community detection algorithm to divide each network into communities, and analyze how these change over time. Since we have a ground truth classification of the sector memberships of the various companies, we can also quantify how well these communities reflect the sector structure.

A popular method to detect communities is to attempt to maximize the modularity of the network. This is an NP-hard problem [36] and so various approximate methods have been proposed, including a spectral method [184] or the Louvain algorithm [25]. These methods have been applied previously to detect communities in financial networks constructed from stock data [205] [113].

The classic formulation of modularity for a network with adjacency matrix A and a vector of community assignments $\boldsymbol{c}$ is [184]

$$Q = \frac{1}{m}\sum_i\sum_j(A_{ij} - \frac{k_i k_j}{m})\delta(c_i, c_j) \tag{5.12}$$

where $m = \sum_i \sum_j A_{ij}$, $\delta(c_i, c_j)$ is the Dirac function, equaling 1 when $c_i = c_j$ (i.e. nodes i and j are in the same community and 0 otherwise) and k_i is the sum of weights of a node. However this is not appropriate when we are looking at graphs with negative edges. Here we use a definition designed for the presence of negative edges. Proposed by Gomez et al [93] we divide the network into positive edges (signified by a $+$) and negative edges (signified by a $-$)

$$A_{ij} = A^+_{ij} - A^-_{ij} \tag{5.13}$$

where

$$A^+_{ij} = \max(0, A_{ij}) \tag{5.14}$$

$$A^-_{ij} = \max(0, -A_{ij}) \tag{5.15}$$

and so the definitions of modularity are

$$Q^+ = \frac{1}{m^+}\sum_i\sum_j(A^+_{ij} - \frac{k^+_i k^+_j}{m^+})\delta(c_i, c_j) \tag{5.16}$$

$$Q^- = \frac{1}{m^-}\sum_i\sum_j(A^-_{ij} - \frac{k^-_i k^-_j}{m^-})\delta(c_i, c_j) \tag{5.17}$$

where

$$m^+ = \sum_i\sum_j A^+_{ij} \tag{5.18}$$

$$m^- = \sum_i\sum_j A^-_{ij} \tag{5.19}$$

$$k^+_i = \sum_j A^+_{ij} \tag{5.20}$$

$$k^-_i = \sum_j A^-_{ij} \tag{5.21}$$

Total modularity is then a scaled version of these

$$Q = \frac{m^+}{m^+ + m^-}Q^+ - \frac{m^-}{m^+ + m^-}Q^- \tag{5.22}$$

We choose the Louvain method [25] to maximize modularity. This method works by maximizing modularity in a greedy bottom up manner. All nodes are initialized into their own random community. We then check the gain in modularity from moving node i from community a to community b. Doing so for all communities, we put node i in the community that maximizes the modularity gain the most, assuming the gain is positive. We then continue to do this over all nodes until we cannot put a node in another community and make a positive gain in modularity. Phase 1 of the algorithm is now complete. In phase 2 each community can then be treated as a node and the edges out to other communities collapsed into one edge per community. The algorithm can then be run again until we do not make a gain in modularity by collapsing the communities.

The gain in modularity from moving isolated node i into a community can be calculated by separately considering the positive and negative edges as follows

$$\delta Q^+ = \frac{\sum_{\text{in}}^+ + 2k_i^+}{m^+} - (\frac{\sum_{\text{tot}}^+ + k_i^+}{m^+})^2 - (\frac{\sum_{\text{in}}^+}{m^+} - (\frac{\sum_{\text{tot}}^+}{m^+})^2 - (\frac{k_i^+}{m^+})^2) \tag{5.23}$$

$$\delta Q^- = \frac{\sum_{\text{in}}^- + 2k_i^-}{m^-} - (\frac{\sum_{\text{tot}}^- + k_i^-}{m^-})^2 - (\frac{\sum_{\text{in}}^-}{m^-} - (\frac{\sum_{\text{tot}}^-}{m^-})^2 - (\frac{k_i^-}{m^-})^2) \tag{5.24}$$

where $\sum_{\text{in}}$ is the sum of weights of all the edges inside the community node i is being moved into, $\sum_{\text{tot}}$ is the sum of weights of the edges to the community. The gains are then scaled by the total weight of positive and negative edges in the graph and combined together

$$\delta Q = \frac{m^-}{m^+ + m^-}\delta Q^+ - \frac{m^-}{m^+ + m^-}\delta Q^- \tag{5.25}$$

A notable advantage of this algorithm is that we do not need to choose the number of communities. Since this algorithm is greedy, we randomise the order of the nodes each run through in phase 1. This of course means that we will achieve different results every time we run the algorithm. Therefore we run the algorithm 10 times on each network to get a mean and standard deviation for any measures taken.

To evaluate the success of the community detection we use the Adjusted Rand Index. Given a set of elements S, and two partitions of this X and Y divided into subsets the Rand Index [215] is defined as

$$R = \frac{a + b}{a + b + c + d} \tag{5.26}$$

where a is the number of pairs of items in the same subset in X and Y, b is the number of pairs of items that are in different subsets in X and Y, c is the number of pairs of items in the same subset in X and a different subset in Y and d in the number of pairs of items in the same subset in Y and different subsets in X. Basically it is very similar to a measure of accuracy if we have a ground truth labelling.

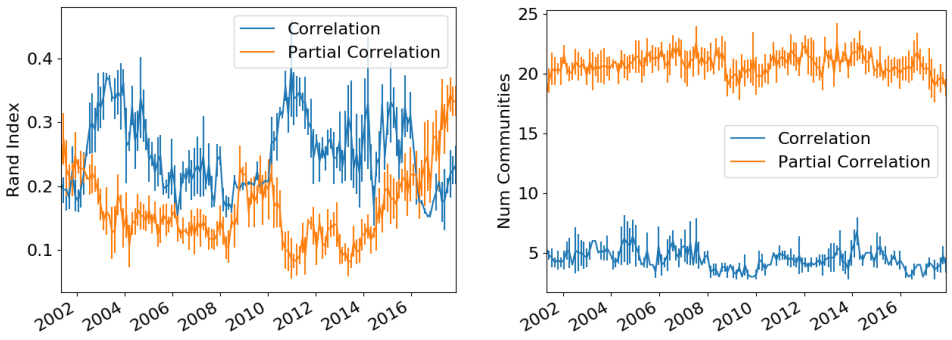
In the Adjusted Rand Index (ARI) a correction is made for chance, using the expected similarity for community detection under a random model:

$$ARI = \frac{R - E[R]}{\max(R) - E[R]} \tag{5.27}$$

Figure 5.11a shows the ARI over time, exhibiting how the networks reflect the known sector structure. The partial correlation networks have less success overall in discovering the sector structure, although both networks show large variations over the time period. We suggest this is to due to the reduction in indirect correlations that the partial correlation coefficient provides - we would expect this to reduce intra-sector correlation strengths which would lead to a reduced success in recovering the sector structure. It is also noticeable that times of disruption seem to lower the ARI for the correlation networks. This could be due to the increased amount of correlation and volatility causing companies to behave more similarly, reducing the ability of the algorithm to separate them [211]. However for the partial correlation networks the ARI actually increases during these times.

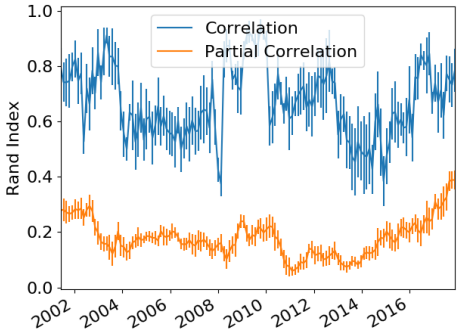
We next look at the number of communities produced. This is shown in Figure 5.11b. The correlation networks have a smaller number of communities than the partial correlation networks, averaging around 4 while the partial correlation networks have an average of around 20. Both choose roughly the same number of communities throughout the entire dataset, although we can see there is a small dip for both networks in 2009. Again this could be due to the increasing correlations during market disruption, making the companies seem more similar. There are 11 actual sectors in the dataset, so neither method is particularly close to the true value.

Finally we study how stable the communities are over time. Since we have 10 partitions per network, we use the adjusted rand index to compare the consistency of the community detection between those from the previous window and those from the next. The results of this are shown in Figure 5.11c. The correlation networks have a much more stable community structure than the partial correlation networks (i.e the ARI between each run is larger meaning more companies are in the same cluster), although the variance does seem to be higher. For the correlation networks there is a large 'break' in 2008 with the consistency dropping considerably, but then both networks have an increase in community stability over the crisis. This is consistent with our finding about network stability in Section 5.4.1, with the partial correlation networks being less stable than the correlation networks.



(A) Adjusted Rand Index

(B) Number of Communities



(C) Community Consistency

FIGURE 5.11: Measures for the community detection over time. The partial correlation networks in general have less success in uncovering the sector structure, have a larger number of communities and have less consistent communities compared to the correlation networks.

5.5 Conclusion and Future Work

In this chapter we have constructed correlation and partial correlation networks from S&P500 returns data using the Ledoit-Wolf covariance estimator. This is designed to cope having more dimensions than samples and always gives us an invertible covariance matrix, which is required for estimation of the partial correlation matrix. We construct 140 networks using windows of data and contrast the correlation and partial correlation networks produced.

Firstly we compare the edge weights in these networks. The partial correlation network has more negative edges than the correlation network and generally has smaller weights. There is however a clear relationship between the two - several edges with a high correlation also have a high partial correlation. Since partial correlation is designed to reduce the effect of indirect correlations this is something we would expect.

Secondly we use the largest eigenvalue and corresponding eigenvector to measure the intensity and stability of the networks. The largest eigenvalue of the correlation network varies significantly depending on the state of the market at that time while the largest eigenvalue of the partial correlation network remains roughly constant. This shows that the partial correlation networks do not have much change in intensity over the dataset. However, using the difference in the largest eigenvector to measure the stability of the network we find the partial correlation networks are significantly less stable than the correlation networks, perhaps showing why minimum risk portfolios tend to be less stable.

Exploring the mean centrality of various sectors using both degree and eigenvector centrality, we find that in the partial correlation networks the sectors all have relatively similar mean centrality with degree centrality. Furthermore macroeconomic factors do not seem to effect the centrality, with all sectors have a fairly consistent mean degree centrality. This is not the case in the correlation networks, where there is clear variation in the centralities, notably during the financial crisis of 2008/2009. The eigenvector centralities show a very different story, with there being more variation in the partial correlation networks rather than less. Macroeconomic effects are also picked up in both networks here, again with the centralities of the sectors moving together during the financial crisis.

Utilizing these networks for portfolio selection, we find there is positive correlation between the centrality of a company and its out of sample Sharpe ratio but there is negative correlation between its centrality and risk. This result is statistically significant and is relevant for both the correlation and partial correlation networks.

Finally we run an altered Louvain community detection algorithm using a version of modularity that is designed for networks with negative edges to attempt to discover whether the sector assignments are replicated in the actual data. We find that the partial correlation networks are less successful than the correlation networks in uncovering these sectors. The correlation networks also produce more stable communities with a lower number of communities than the partial correlation networks. This indicates that in general the partial correlation networks have less stable structure than a correlation network constructed from the same data.

Since we have struggled to extract easily explainable structure from these partial correlation networks (when compared to the Pearson correlation ones) we would not recommend their use in any applications interested in global properties of the networks. We have however not sufficiently explored how individual nodes may be differently expressed in these networks. This could be the topic of a future study.

Chapter 6

Quantifying Influence in Financial Markets via Sparse Partial Correlation Estimation

6.1 Introduction

In this chapter we propose to use the SPACE [202] (Sparse PArtial Correlation Estimation) to estimate a partial correlation network from financial returns data. Justification for why we wish to do this can be found in Chapter 5. Here we briefly re-explain how the SPACE algorithm works (a larger review on the estimation of partial correlation matrices is available in Chapter 3).

Given an n (number of samples) by p (number of companies) dataset X, we note we can recover row i (not including the diagonal) of the precision matrix by regressing variable i upon the rest.

$$\arg\min_{\boldsymbol{\beta_i}} ||\boldsymbol{x_i} - X_{-i}\boldsymbol{\beta_i}||_2^2 \tag{6.1}$$

The solution of this is $\beta_{ij} = -\frac{\Theta ij}{\Theta_{jj}}$. We can now get the partial correlation matrix (P) using

$$P_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}\Theta_{jj}}} \tag{6.2}$$

We can note from these that $\beta_{ij} = P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}$

Meinshausen and Bulhmann [170] propose neighbourhood selection where we solve (6.1) using the lasso. However neighbourhood selection does not take advantage of the symmetry of the problem (the precision matrix is symmetric) and the imposition of sparsity is row-wise rather than over the entire matrix. Since many real-world networks are believed to have hubs (nodes with a large number of connections) we may wish to impose sparsity on the entire network rather than on rows. With SPACE we formulate 1 large problem rather than p smaller ones and impose sparsity on the partial correlation matrix rather than the precision matrix, and so the objective function becomes

$$L(P, \Theta) = \frac{1}{2}\sum_{i=1}^{p} ||\boldsymbol{x_i} - \sum_{j\neq i}^{p} P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}||_2^2 + \lambda \sum_{1\leq 1<j\leq p} |P_{ij}| \tag{6.3}$$

We alternate between solving for P and for the diagonal of Θ until the problem converges - this typically takes around 2-3 iterations. To solve this we formulate a large lasso problem - on the order of np by $p(p-1)/2$. This is done by creating 2 new

regression variables $\hat{Y}$ and $\hat{X}$. $\hat{Y}$ is a np column vector containing all the variables stacked upon each other

$$\hat{Y} = (\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_p})^T \tag{6.4}$$

Next we need to create a matrix that contains the predictors for the appropriate $\hat{Y}$ values.

$$\hat{X} = (\bar{X}_{(1,2)}, \dots, \bar{X}_{(p-1,p)}) \tag{6.5}$$

$$\bar{X}_{(i,j)} = (0, \dots, 0, \sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}^T, 0, \dots, \sqrt{\frac{\Theta_{ii}}{\Theta_{jj}}}\boldsymbol{x_i}^T, 0, \dots, 0) \tag{6.6}$$

The following problem is then solved to calculate the off-diagonal values of the partial correlation matrix

$$\min_{\boldsymbol{a}} ||\hat{Y} - \hat{X}\boldsymbol{a}|| + \lambda||\boldsymbol{a}||_1 \tag{6.7}$$

where $\boldsymbol{a}$ contains the off-diagonal values. Once we have run this we can estimate the diagonal as following

$$\frac{1}{\Theta_{ii}} = \frac{1}{n}||\boldsymbol{x_i} - \sum_{j \neq i}^{p} P_{ij}\sqrt{\frac{\Theta_{jj}}{\Theta_{ii}}}\boldsymbol{x_j}||_2^2 \tag{6.8}$$

Similarly to chapter 5, our goal is again to extract some form of stable, interpretable structure from the networks. We firstly hope there will be some structure which persists over time in the networks. We also hope to be able to see the pre-known sector structure in these networks.

6.2 Methods and Data

Since financial data is non stationary we use a window of 300 days and slide along this 30 days at a time to obtain a sample where the data is more stationary. The data inside each window is then normalized to have a mean of 0 and a variance of 1 using the z-score. This is mostly for the benefit of the SPACE algorithm, as it requires standardized data. We use the SPACE method to estimate a sparse partial correlation matrix, using a Bayesian Information Criterion (BIC) method to select λ. To do this we run the SPACE algorithm for 50 linearly spaced values of λ from 1 to 200 initially. If either of the limits are selected we increase or decrease the limits λ as required and rerun the procedure. The network with the lowest BIC is selected. The diagonal is filled with zeros to avoid self links and we use this as an adjacency matrix to construct a network. We then study how the networks change over time.

For our study we use daily log returns from the S&P500. If there is less than 10% of the data missing for a particular stock we fill it with the price from the previous day, or if the data is missing from the start, from the first day when the stock is traded. If there is more than 10% missing we discard the data for that stock. We use the closing value on the day, from 01/03/2000 to 05/12/2017. Overall we had 4510 days of return data for 345 stocks, and this was divided into 140 windows.

We make use of Python, NumPy and SciPy for general scripting, pandas for handing the data, sklearn, Networkx and gephi for the network analysis. If the reader is interested in reproducing our work, the code and data can be found at https://github.com/shazzzm/financial_network_inference

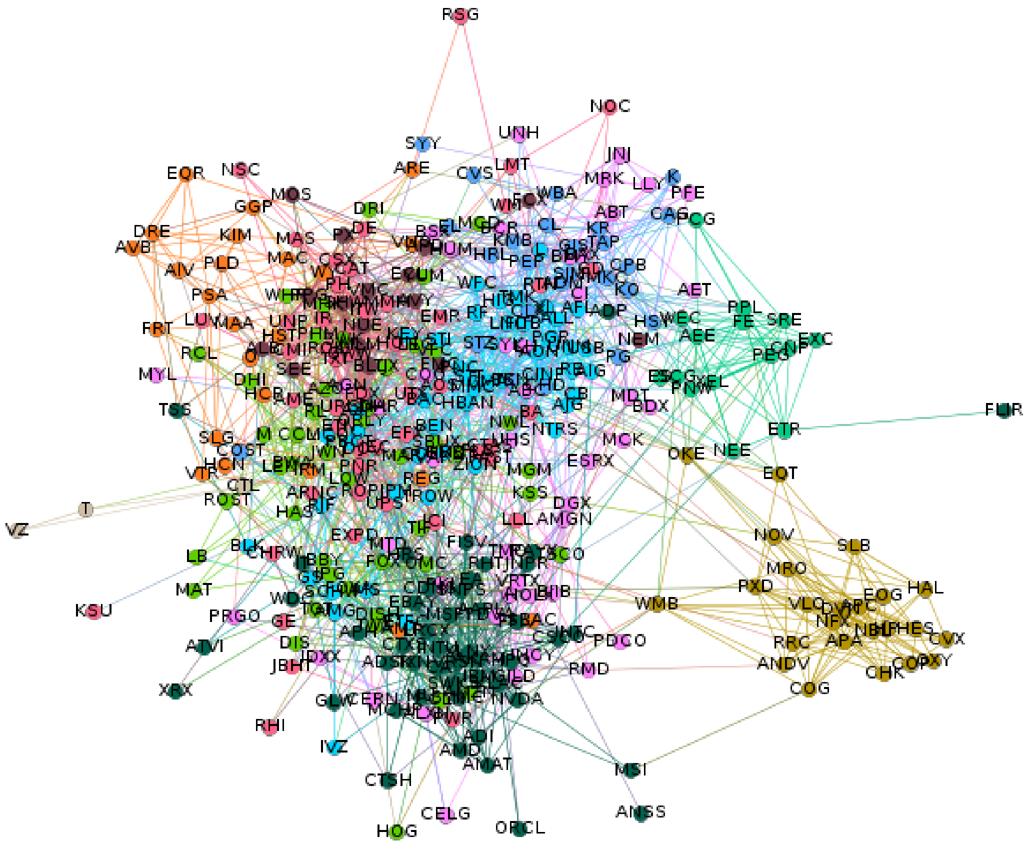


FIGURE 6.1: Example of a generated financial network. Nodes are coloured according to the sector they belong to. We can see that clusters of companies all in the same sector tend to cluster. This seems intuitive - companies that have lines of business that are related are likely to be affected by similar underlying factors, ensuring their stocks are correlated

6.3 Results

Firstly we show an example network in Figure 6.1. This is the network inferred from the first window. The nodes are coloured according to the sector the company belongs to. From this we see that companies in the same sector tend to cluster together, with a noticeable group of companies in the energy sector in the bottom right. Since larger stock movements tend to be affected by macroeconomic factors, we would expect that companies that have similar lines of business would be affected by these factors in the same way and so would have correlated stock price movements.

To further study this sector clustering we sum up the number of connections between each sector and the others (including themselves) over the entire set of networks and plot the results as a heat-map, shown in Figure 6.2. As the diagonal component tends to dominate (i.e. companies in a sector are mostly connected to those in the same sector) we take the logarithm of the values to make the heatmap easier to understand. The weight of an edge is not taken into account here, only its presence. Firstly we notice that the diagonal contains the largest weights - i.e. companies are most likely to be connected to those in the same sector. Secondly, the energy sector mostly seems to connect to itself and has relatively few outside

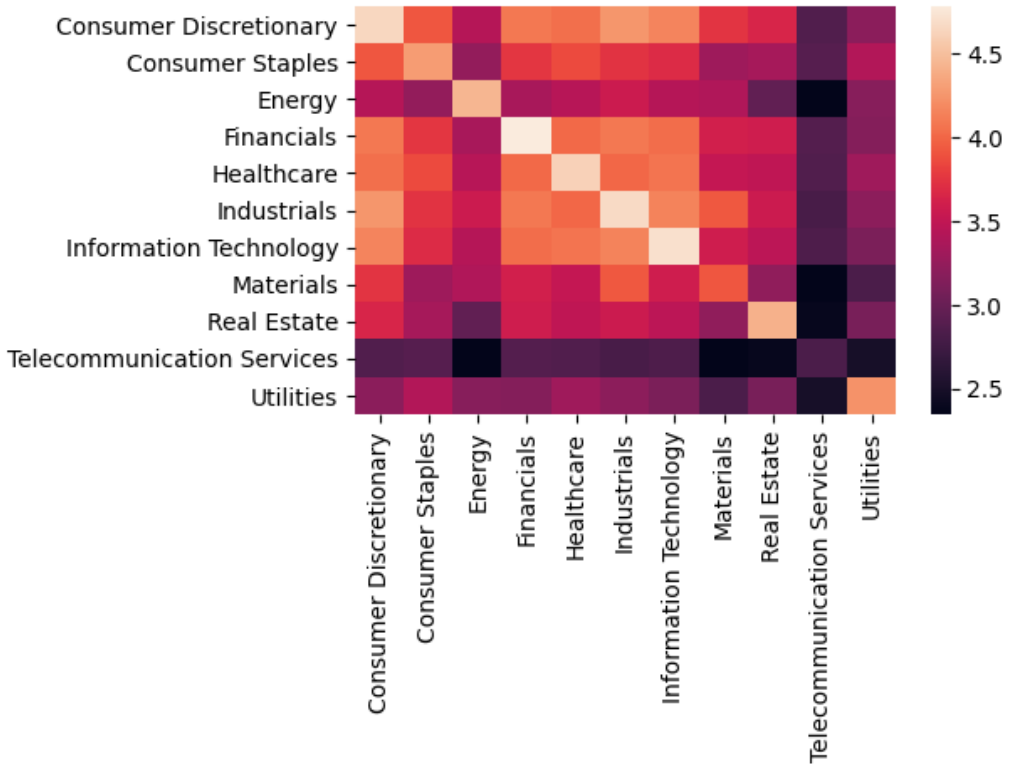


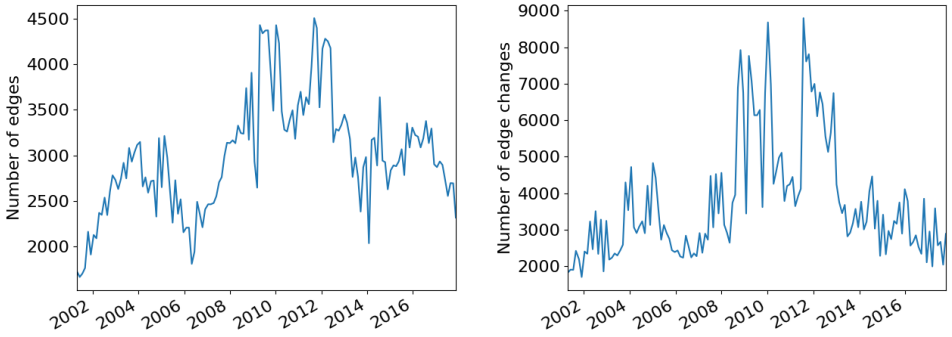
FIGURE 6.2: Log of the number of connections (i.e. weights are not taken into account) between each sector and the others. Lighter colours indicate a larger number. The lighter diagonal indicates that companies mostly seem to be connected to those in the same sector, and we note the energy sector seems to be mostly self contained while the financial, heath-care, industrial, technology and consumer discretionary sectors make up most of the interactions

connections. Thirdly we notice there is a 'block' of sectors that seem to be interconnected - the financial, health-care, industrial, technology and consumer discretionary sectors. Meanwhile the telecommunications sector has very few links, indicating its relative unimportance.

Next we see how the number of edges in the networks change over time. A plot of this is shown in Figure 6.3a. Generally we can see that during times of market growth and calmer conditions there is a smaller number of edges. However during times of market disruption and crashes the number of edges rapidly increases. We do know from previous work that the market tends to become more correlated during times of market disruption [211] and that the minimum spanning tree tends to shrink [195]. Therefore it would make sense that the number of edges increases during these times as the amount of correlation in general has increased.

To characterize the stability of the networks we look at the number of the edges that change between networks inferred from adjacent time periods. A plot of this is shown in Figure 6.3b. Firstly we observe the networks are not very stable - most of the edges change between adjacent time periods. The networks also change much more in times of market instability.

Next we study how the influence of various sectors changes over time. To measure this we study the centrality of the companies in the sectors and take the mean centrality for each sector. The presence of negative edges in the networks makes this more



(A) Number of edges (B) Number of edge changes

FIGURE 6.3: Number of edges and edge changes in each network as we move through time. Generally we can see that during times of market growth we have a smaller number of edges, while during times of significant market movement and change we tend to get far more edges. Furthermore there are more edge changes during times of stress.

complicated as this means that many usual measures of centrality are no longer guaranteed to be positive. To solve this we use the sum of the absolute values of edges of a node to measure its centrality, which will always give us a positive number and we divide it by the total sum to ensure the sum of all centralities is 1. We show how these vary in Figure 6.4.

From the start of the dataset we can see that the financial sector has the highest centrality for most of the dataset. However between 2009 and 2013 this is not the case, with the consumer discretionary, industrial and technology sector also briefly holding the top position. After 2008 we can see the centrality spread decreases, particularly around 2009 where several groups of sectors seem to cluster together for a year long time period before starting to diverge again. At this point the financial sector regains its position as the most central. As previously mentioned the number of edges in the network increases during times of market disruption and suddenly companies previously unconnected become connected. This could be due to the increased correlation that occurs during times of market distress [211]. This would explain why the centralities move closer together - suddenly there are more edges in the network sharing the weights out further. For most of the dataset we can also see that telecommunications, consumer staples and utilities have a low centrality and generally do not interact much with the rest of the sectors. This is reflected in previous work [126] [95], and our heat-map in Figure 6.2. Generally these sectors are regarded as less influential in the US economy.

Finally we are interested in exploring the consequences of this work for portfolio selection. Previous work in this field has suggested selecting companies that have a low centrality measure [203] [210], but this is in relation to correlation networks. We study here how this advice transfers over to our partial correlation networks. Our first remark is to draw attention back to Figure 6.3b, where we measure the number of edge changes as a proxy for measuring the stability of the networks. It has been noted that portfolios constructed using the empirical covariance estimator tend to be unstable and have wild swings in weights [63]. Even for our regularized estimator we can see how this would be the case - if the precision matrix changes rapidly we would expect a portfolio calculated using it to also change rapidly. Since we have this relationship between portfolio optimization and the centrality of a

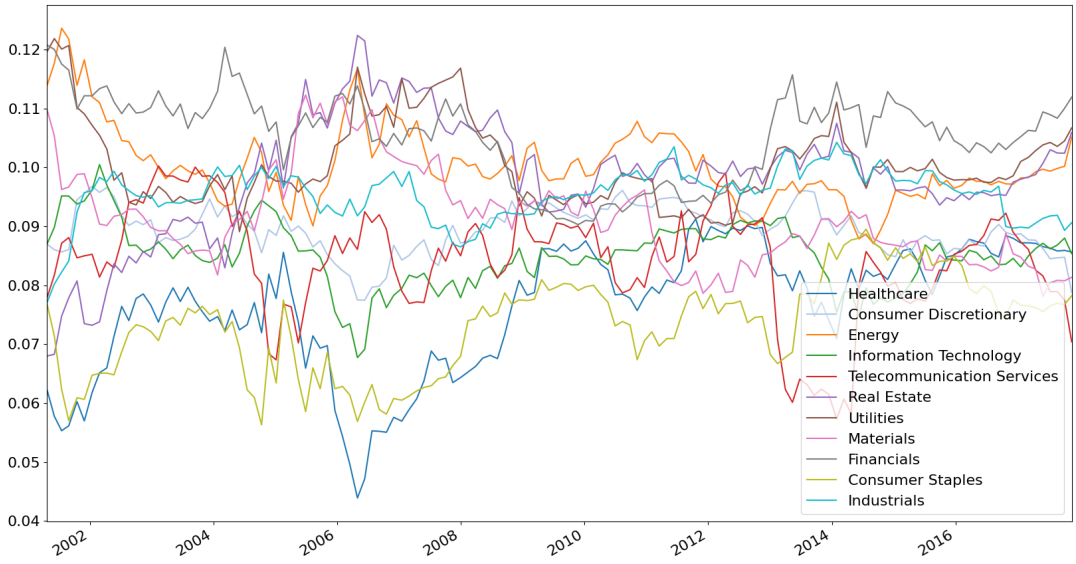


FIGURE 6.4: Plot of how the centrality of each sector varies over time. We sum the absolute value of each edge to measure centrality as the nodes may also have negative edges and divide by the sum to ensure we have a centrality that adds to 1. In general we note the financial, consumer discretionary, industrial and technology sectors are important in the US markets, while the telecommunications, materials and utilities have a low centrality and interact less. These results are reflected in previous work [126] [95]. We can also see the effects of the 2009 market crash, with the centralities clustering together more. This could be due to the increase in correlations in the market sharing the centrality out more.

company, we investigate how this presents itself in our dataset. We infer networks from the 300 day time period and then evaluate the mean return (μ) and risk (we use standard deviation, σ) of each stock on the next window, and calculate the Sharpe ratio (defined as $\frac{\mu}{\sigma}$). Due to its robustness, we use the Spearman correlation coefficient to measure the relationship with a two sided test for significance, with a null hypothesis that the datasets are uncorrelated. The Spearman correlation is 0.0965, indicating there is relatively little relationship here. However the Spearman correlation between the centrality and the out-of-sample risk is -0.167 with a p-value of less than 0.0001 indicating the result is significant. To explore how this changes over time we plot the Spearman correlation between the centrality of a company and its risk in the next window in Figure 6.5. In general there seems to be a negative correlation between the two, except from 2009 - 2013 where the relationship becomes much less significant. It is also at these points where the p-values indicate the relationship is not significant, although due to space we omit a plot of these. This also indicates it is challenging to create low risk portfolios during periods of market disruption.

6.4 Conclusions

In this chapter we have inferred partial correlation networks from S&P500 daily returns using the SPACE algorithm, analyzed how these networks vary over time and how the importance and influence of various sectors changes. We have found that the number of edges increases during times of market disruption, possibly due

FIGURE 6.5: Centrality of a company compared to its risk on the next window of data (i.e. out-of-sample). In general there seems to be negative correlation between the two aside from during 2009-2013, although at this point we also fail to reject the null hypothesis that the data is not correlated

to the increased amount of correlation during these times. We find that the most central sector in the market is usually the financial sector, but this mantel is also taken by the technology, industrial and consumer discretionary sectors. Companies tend to have more connections with those in the same sector, but certain sectors tend to have more connections across the market (e.g. financial, consumer discretionary) and others (e.g. energy) tend to be more inward looking. Finally we study the relationship between centrality in the network and the out-of-sample Sharpe ratio. We do not find a significant relationship between the out-of-sample Sharpe ratio and the centrality of a company, but we do find a mild negative relationship between centrality and out-of-sample risk.

While there is evidence of some form of sector structure in these networks, the networks themselves are very unstable. This is a large issue if we desire to use the networks for financial applications, as it indicates many of the edges present are noise. Therefore, as with the dense networks, we would not recommend using them in their current form.

Chapter 7

Construction of Minimum Spanning Trees using Rank Correlation

7.1 Introduction

Most of the literature studying correlations between asset returns tends to use the Pearson correlation coefficient. While the interpretability of the Pearson correlation coefficient is a big plus, it assumes normality, something which most assets return distributions do not follow [59], and is sensitive to outliers. There are of course correlation measures that do not suffer from these issues, namely those based on rank. Rank correlation methods calculate the correlation between the ranks of variables, which tends to remove the effects of outliers while still giving a measure of the degree to which two variables increase or decrease together. However to the best of our knowledge, most of the literature which studies the correlations between asset returns tends to use the Pearson correlation coefficient. Therefore in this chapter we compare networks inferred from stock returns using Pearson, Spearman and Kendall's τ correlation in order to see if the robustness of these rank correlations can give an improved picture of the stock markets.

To evaluate the rank MSTs, we compare them to MSTs constructed using Pearson correlation from the same data. Broadly speaking, we expect that the rank MSTs should be more stable than those constructed from Pearson correlation, and maintain their structures for longer periods of time. If non-linearities and/or outliers are also affecting the Pearson MSTs significantly, we would expect greater differences between the Pearson and Rank MSTs during times of market volatility too. We therefore use a set of network measures which quantify these properties, which are defined later on in the chapter.

The Pearson correlation between two variables r_i and r_j is defined as follows

$$C_{ij} = \frac{\sum_{i=1}^{n}(r_i(t) - \bar{r}_i)(r_j(t) - \bar{r}_j)}{\sqrt{\sum_{i=1}^{n}((r_i(t) - \bar{r}_i)^2(r_j(t) - \bar{r}_j)^2)}} \tag{7.1}$$

To calculate the Spearman correlation, we firstly sort the values, replace each value with its rank, and calculate the Pearson correlation between the ranks. This then measures the degree to which two variables monotonically increase or decrease together. Kendall's τ is slightly more complicated, measuring the relationship by considering the number of concordant pairs vs the number of discordant pairs. A pair of observations $(r_i(t), r_j(t)), (r_i(t+1), r_j(t+1))$ is concordant if $r_i(t) > r_j(t)$ and $r_i(t+1) > r_j(t+1)$ or if $r_i(t) < r_j(t)$ and $r_i(t+1) < r_j(t+1)$. It is discordant

if $r_i(t) > r_j(t)$ and $r_i(t+1) < r_j(t+1)$ or if $r_i(t) < r_j(t)$ and $r_i(t+1) > r_j(t+1)$. The τ-a formula simply counts the number of concordant pairs vs the number of discordant pairs, divided by the total number of pairs. This does not however take into account any ties that might occur in the data, so we use the τ-b formulation, defined as

$$\tau = \frac{n_c - n_d}{\sqrt{(n_0 - n_1)(n_0 - n_2)}} \tag{7.2}$$

where n_c is the number of concordant pairs, n_d is the number of discordant pairs, $n_0 = \frac{n(n-1)}{2}$, $n_1 = \sum t_a(t_a - 1)/2$, $n_2 = \sum u_a(u_a - 1)/2$, t_a is the number of values in the ath group of ties for variable i, u_a is the number of values in the ath group of ties for variable j. Any reference to τ in the rest of the chapter refers to this τ-b formation.

7.2 Data and Software

The data we use is downloaded from Yahoo Finance. For the UK data we use the FTSE100 companies, for the US returns we use the S&P500 companies and for the German data we use the DAX30 companies. We use returns from 2000/03/01 to 2019/10/21. For each dataset any company missing more than 10% of its data is removed, and any missing values are filled forwards. If the values are missing from the start we backfill from the first good value. From this data we take a window of 504 days and slide along 30 days at a time. This leaves us with 4789 days of returns data from 229 companies for the US, 5063 days of returns data from 70 companies for the UK and 5066 days for 23 companies for Germany. Germany will also be referred to as DE in the rest of this chapter. Each company is tagged with a sector from the GICS classification using information from Bloomberg. This places each company into 1 of 11 sectors, Technology, Real Estate, Materials, Communications, Energy, Financials, Utilities, Industrials, Consumer Discretionary, Healthcare or Consumer Staples.

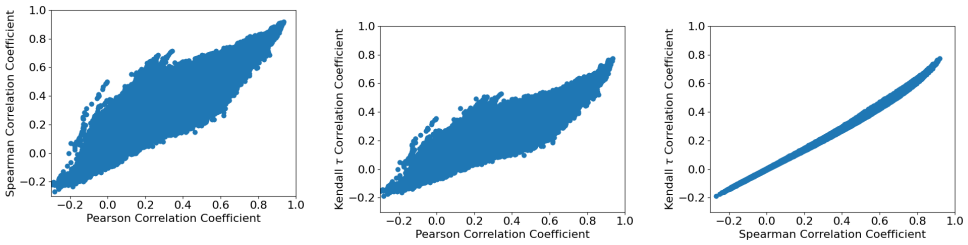
We make use of Python, NumPy and SciPy [193] for general scripting, pandas [169] for handing the data, statsmodels [227] for some of the statistical analysis, matplotlib [110] for plotting, arch [233] for the implementation of the circular bootstrap, TopCorr for the MST construction, Networkx [96] for the network analysis and Cytoscape [229] for the MST visualization. The code and data is available at `https://github.com/shazzzm/rank_correlation_msts`.

7.3 Results and Analysis

7.3.1 Correlation Matrix Analysis

Firstly we analyze the full correlation matrices with no filtration. A starting point is to look at the correlation coefficient for the same set of values. Figure 7.1 shows a set of scatter plots comparing the correlations. From this we can see there is a degree of agreement between all, and generally larger correlations are more likely to be similar. However there is a 'fat' middle when comparing the rank correlations to the Pearson correlation, where there can be significant disagreement. Spearman and τ seem to be very similar, with there being a strong relationship between the two, although the τ correlations are slightly smaller than the Spearman ones.

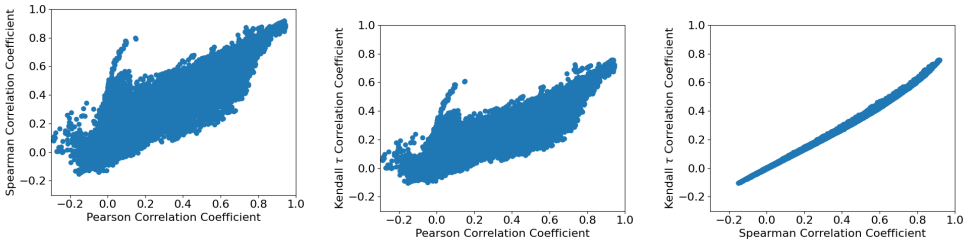
The largest eigenvalue of the correlation matrix is a measure of the intensity of the correlation present, and in matrices inferred from financial returns tends to



(A) Pearson vs Spearman - US

(B) Pearson vs τ - US

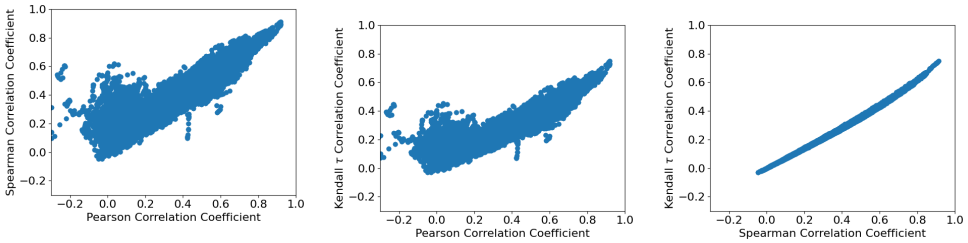
(C) Spearman vs τ - US



(D) Pearson vs Spearman - UK

(E) Pearson vs τ - UK

(F) Spearman vs τ - UK

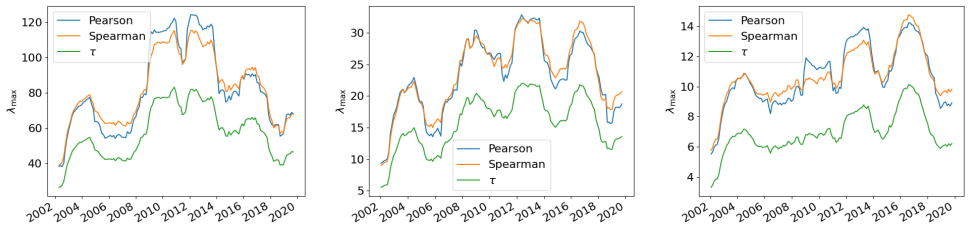


(G) Pearson vs Spearman-DE

(H) Pearson vs τ - DE

(I) Spearman vs τ - DE

FIGURE 7.1: Relationship of the correlation coefficients for each country across the entirety of each dataset. There is a large degree of agreement and larger correlations are more likely to be similar, but the 'fat' middle is notable when comparing the Pearson and rank correlations. The rank correlations themselves are relatively similar, with the τ correlations being smaller than the Spearman correlations.



(A) US (B) UK (C) DE

FIGURE 7.2: Largest eigenvalue (λ_{max}) in the networks over time. From this we can see the Spearman correlation has a slightly smaller largest eigenvalue than the Pearson correlation, while the τ correlation is much smaller. The rank methods also have a smaller range than the Pearson correlation. The volatility of the markets at times of stress is likely to lead to more outliers, so the robustness of the rank correlations to these could be causing the reduced variance of the largest eigenvalue.

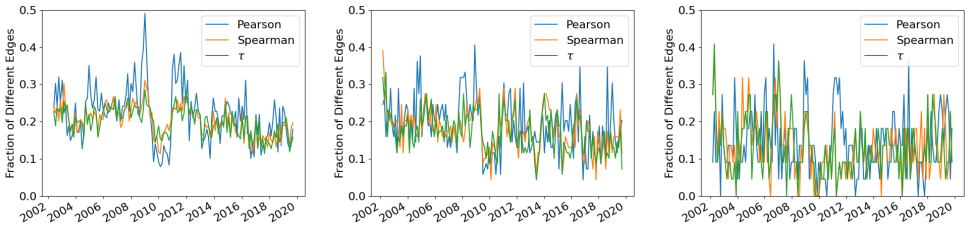
be significantly larger than the second largest [139] [206]. Generally this largest eigenvalue is larger during times of stress and smaller during times of calm [72] [139]. Firstly we study how this varies over time for each correlation measure. This is shown in Figure 7.2. For all of the networks there is a similar shape, with it peaking during times of market stress and dropping during times of calm. The Spearman and Pearson correlations have relatively similar values, although the Spearman has a smaller range. The τ correlation is much smaller than the other two at all times, and also has a smaller range. Times of stress and volatility tend to bring more outliers in returns data, which could be the cause of the difference in largest eigenvalue. Figure 7.1 has shown us that the τ correlation coefficients tend to be smaller than the others, which could explain why the largest eigenvalue is consistently smaller too.

7.3.2 MST Stability

In this section we analyze and compare the stability of the trees constructed using the various coefficients.

Firstly we focus on measuring the fraction of edge changes between MSTs adjacent in time, which quantifies how stable the trees are, and how well change in the market is detected. The results are plotted in Figure 7.3. For the US and the UK, the Pearson MSTs tend to have much greater spikes in edge changes compared to the rank MSTs, indicating the rank MSTs are more robust to the outliers present during times of stress. Interestingly the number of edge changes seems to drop during the financial crisis in 2009 for all countries. Other authors have noted that by some measures the markets could be considered more stable during these times [72] [134], but we have not found that this has been mentioned in the context of MSTs. For Germany the results are harder to interpret, as the German MSTs are much smaller, meaning small changes are much more significant as an overall fraction and so we report the mean and standard deviation (s.d.) of these MST differences. The mean differences are are 0.138 (s.d. 0.089) for the Pearson MST, 0.131 (s.d. 0.080) for the Spearman and 0.128 (s.d. 0.077) for the τ MSTs. This indicates that the rank MSTs do change less than the Pearson ones, but the difference is small.

This therefore shows that the Spearman and τ MSTs tend to be more stable than Pearson MSTs for all of the countries. This is particularly prominent at the start of the financial crisis for the US and the UK, with the Pearson MSTs showing a large spike



(A) US (B) UK (C) DE

FIGURE 7.3: Edge difference between adjacent MSTs. From this we can see the MSTs inferred using rank correlation are far more stable with regards to time than those inferred using Pearson correlation. While all of the trees seem to become more similar during the financial crisis (2009 - 2011), the Pearson MSTs show a big reconfiguration as the crisis starts, while the τ MSTs shows no spike before dropping.

in difference, while the Spearman and τ MSTs show little or no change in difference. In this particular situation we would expect the heavy tails to affect the Pearson correlation between two assets more than the rank methods, and this should change the edges selected by the MST construction procedure.

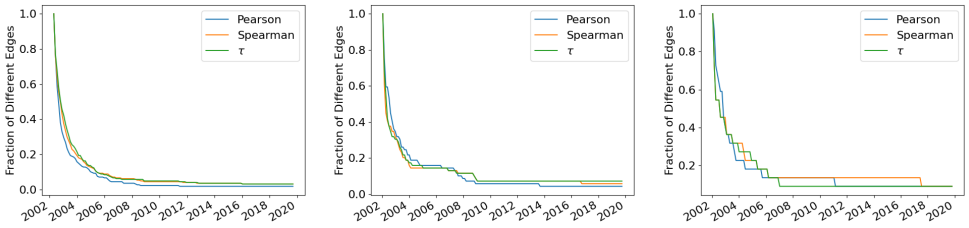
Next we measure how the MSTs have changed from the first inferred tree using the multi step survival ratio [196] - the fraction of edges that have been consistently maintained for each tree from the first to the current. This measures the life of an edge and shows us how the tree evolves. This is defined as

$$\frac{1}{p-1}|E(t) \cap E(t-1) \dots E(t-k+1) \cap E(t-k)| \tag{7.3}$$

where $E(t)$ is the edge set at that moment in time, k goes from 1 to $t-1$ and $|S|$ is the cardinality of the set S. A plot of this is shown in Figure 7.4. From this we can see that quite rapidly the trees differ from the original for all countries, with 70% of the edges changing within 2 years. For our experiments, what is particularly interesting is that for the US and the UK the rank MSTs maintain slightly more edges than the Pearson MSTs, but the difference between the τ and Spearman MSTs is very small. For Germany, the Spearman MSTs actually maintain the most edges for the longest time period, followed by the Pearson MSTs and then the τ MSTs.

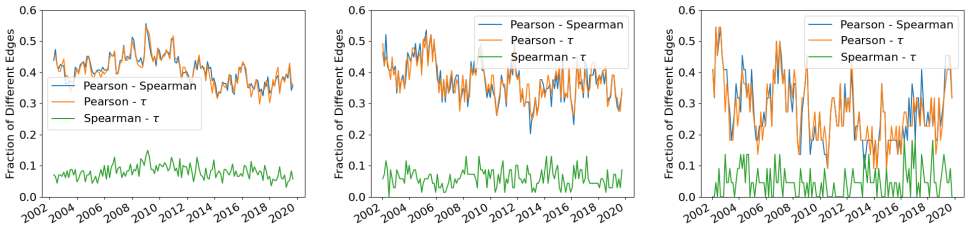
Over the entire dataset for the US, the Pearson MSTs maintain 4 edges, the Spearman MSTs 7 edges and the τ MSTs 8 edges. For the UK the Pearson MSTs maintain 3 edges, the Spearman MSTs 5 edges and the τ MSTs 4 edges. For Germany all three maintain 2 edges. All of these edges are intrasector aside from one in the German τ MSTs, which is BASF to Bayer (Materials to Healthcare).

There is of course the question of how the difference between the MSTs changes over time. We measure the fraction of edges that differ between the three MSTs and plot it in Figure 7.5. From this we can see there is a significant difference in the presence of edges between the rank MSTs and the Pearson MSTs for all countries. The difference does seem to increase during the banking and financial crises, with notable peaks occurring during 2008 and 2009. There seems to be relatively little difference between the two rank MSTs, with less than 10% of the edges being different for every country for most of the time period. This difference between the rank methods does not seem to be particularly affected by market conditions.



(A) US (B) UK (C) DE

FIGURE 7.4: Multi-step survival ratio for the MSTs. This gives us a measure of how long the edges persist for. Most of the edges disappear very rapidly, with around 70% changing within 2 years. The rank MSTs seem to be slightly more stable than the Pearson MSTs, maintaining more edges from the initial tree.



(A) US (B) UK (C) DE

FIGURE 7.5: Fraction of the edges that differ between the various MSTs. The US has the largest difference, followed by the UK and then Germany, which indicates size of the MST influences the edge difference. In general for the US and the UK around 40% of the edges differ, although this does increase during the financial crisis, particularly in 2009. Since the German tree is so much smaller there is a larger range in these values, but on average around 30% of the edges different between the Pearson and rank MSTs.

7.3.3 Node and Sector Centrality

These networks can be used to study the importance of economic sectors in the markets of the selected countries, and how they might be implicated in various crises and growth periods that occur. Here we are interested in how sectors may be differently expressed in the MSTs, to see if the different correlation coefficients can give an alternative picture of the market at that moment.

To measure this we look at how the centrality of a sector varies by calculating the mean of the centrality of all the nodes in said sector. This reduces the effect of the different numbers of companies in each sector. Then to make comparisons between the differently sized MSTs easier, we normalize the sector centrality so the sum of all sector centralities in an MST is 1. To express mathematically, we calculate the centrality of sector s from the set of all sectors S as follows

$$\frac{1}{\sum_{j \in S} C_j} \frac{1}{|s|} \sum_{i \in s} c_i \tag{7.4}$$

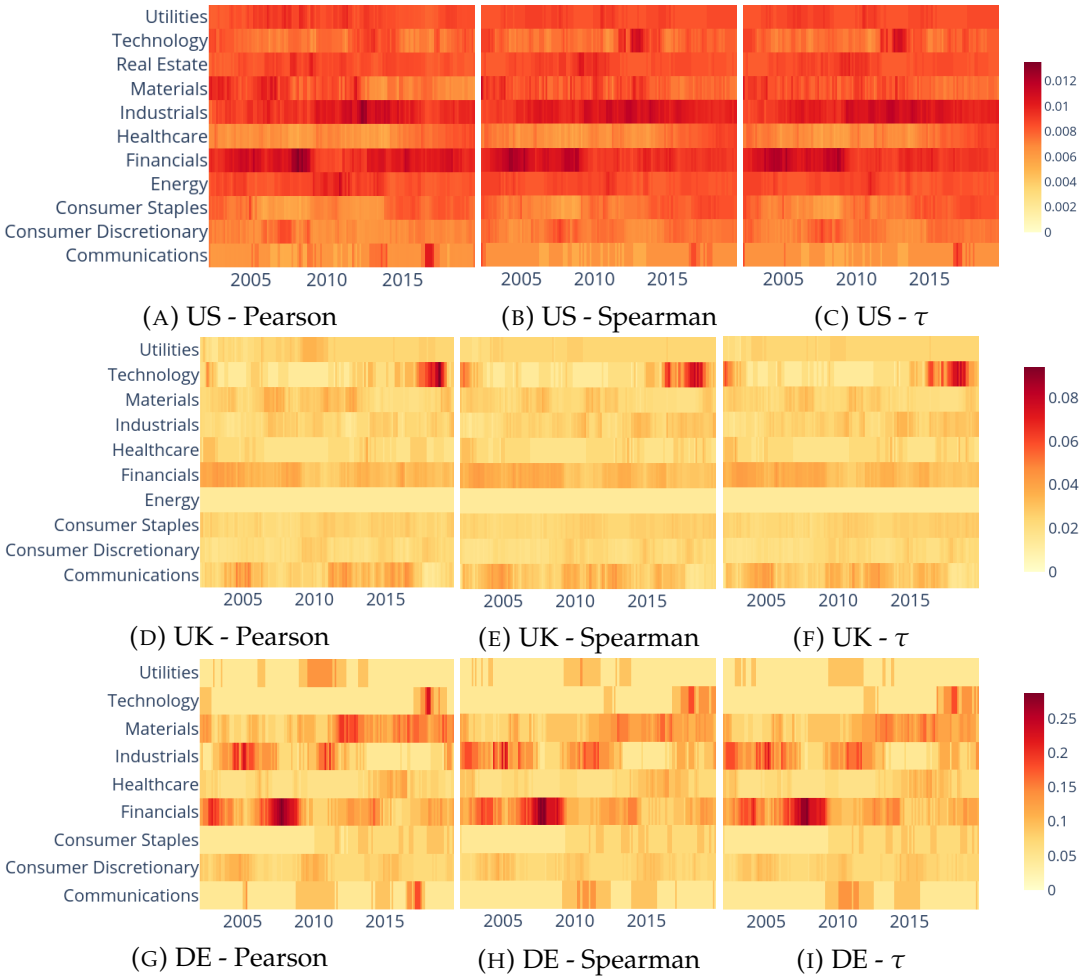
where c_i is the centrality of node i, and $|s|$ is the cardinality of set s. We use both unweighted degree centrality and betweenness centrality to measure the influence of a sector. Betweenness centrality is calculated by looking at the fraction of shortest paths that pass through a node, and allows us to get a different perspective on which nodes are more important. The results are shown in Figures 7.6 (degree centrality) and 7.7 (betweenness centrality).

Firstly we focus on the sector degree centrality. For the US, the Financials and Industrials sectors are important in all three MSTs for the entire dataset. They also seem relatively similarly expressed over time in all three. The Materials sector is regarded as more important at the start of the Pearson MSTs (from 2003 - 2008) but less so in the rank MSTs. The Technology sector becomes more important in the rank MSTs than in the Pearson MSTs, while the Energy sector is slightly more important in the Pearson MSTs from 2009 - 2010 compared to the rank MSTs. The Consumer Discretionary, Healthcare and Utilities sectors all are relatively similar throughout the dataset.

For the UK again the Financials sector is important in all three MSTs. The most notable feature is the large spike in importance of the Technology sector from 2015, which is present in all three MSTs again. The Utilities sector seems more important in the Pearson MSTs than the rank ones, while the Industrials sector seems slightly more important in the rank MSTs, notably from 2015 onwards.

For Germany the Financials, Industrials and Materials sectors are again important in all three MSTs. The Industrials sector seems more important in the rank MSTs, notably from the start of the dataset until 2008 and then from 2016 onwards. The Communications sector is quite differently expressed between the Pearson and rank MSTs, but is very small, so the effects of one company are large. The spike in the centrality of the Technology sector in 2017 is more intense in the Pearson MSTs than the rank ones, but the rest of the centrality seems similar.

In general it seems the degree centrality of most sectors is relatively similar in all three MSTs, with only small differences occurring. Furthermore the sector centralities of the rank MSTs are virtually identical.

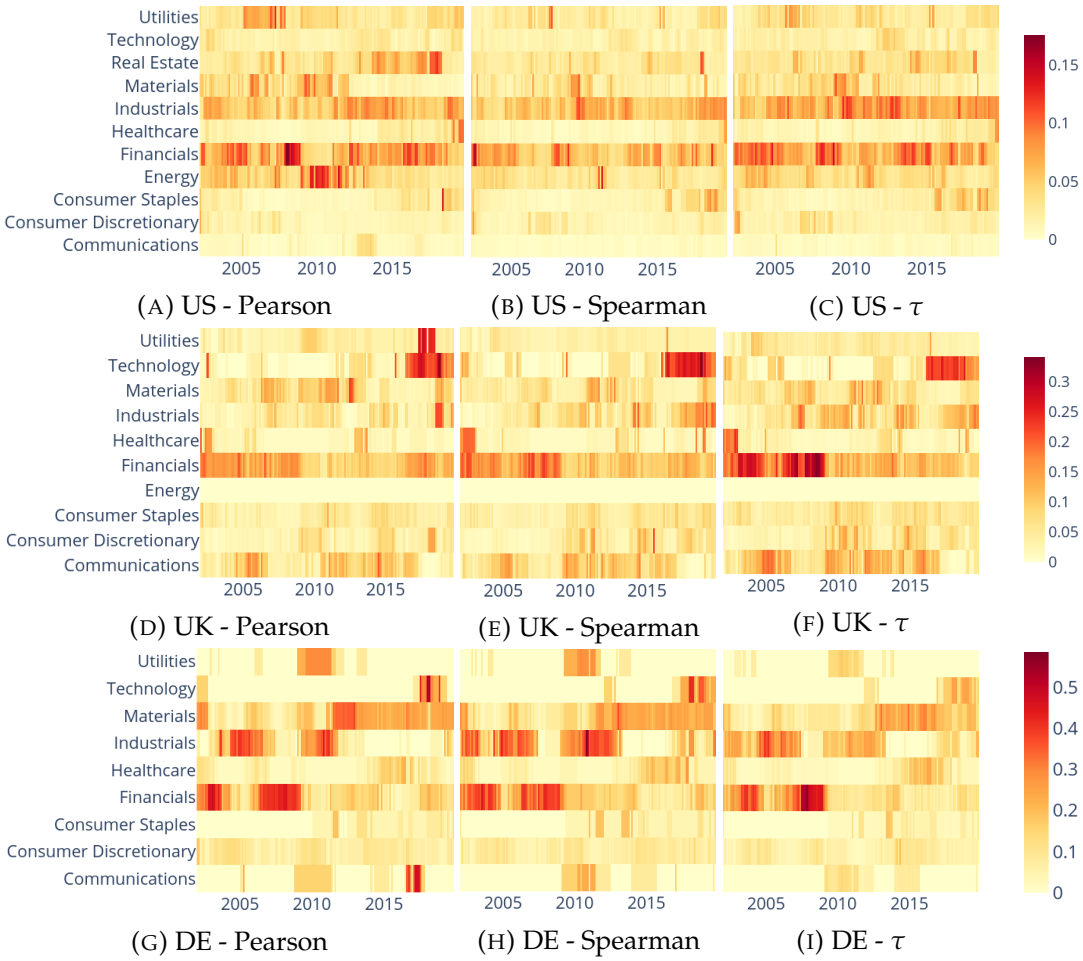


(A) US - Pearson (B) US - Spearman (C) US - τ

(D) UK - Pearson (E) UK - Spearman (F) UK - τ

(G) DE - Pearson (H) DE - Spearman (I) DE - τ

FIGURE 7.6: Degree centrality for each of the sectors over time



(A) US - Pearson (B) US - Spearman (C) US - τ

(D) UK - Pearson (E) UK - Spearman (F) UK - τ

(G) DE - Pearson (H) DE - Spearman (I) DE - τ

FIGURE 7.7: Betweenness centrality for each sector over time

Next we study the betweenness centralities of the sectors. Starting with the US, again the Industrials and Financials sectors are central for the majority of the dataset in all three MSTs. Both seem more central in the τ MSTs than the Pearson or Spearman. The Energy sector becomes notably central in the Pearson MSTs from 2009 - 2012 but this is not reflected in the rank MSTs, and there is also a rise in centrality for the Utilities sector in the Pearson MSTs that again is not seen in the rank MSTs. The Consumer Staples sector seems more central in the rank MSTs, notably towards the end of the dataset from 2015 onwards, as does the Consumer Discretionary sector.

For the UK the Financials sector is again central throughout the dataset in all three MSTs, but has a higher centrality in the τ MSTs. The Technology sector also becomes very central in all three from 2016, but has the largest centrality in the Pearson MSTs. The Utilities sector is not central for most of the dataset in any of the MSTs, but has a peak during 2016 for the Pearson MSTs, but not in the rank MSTs. The Communications and Consumer Staples sectors are relatively similarly expressed across all MSTs, and are not very central. The centrality of the Materials sector varies similarly in all MSTs, but has a higher peak value in the Pearson MSTs, and the Consumer Discretionary is similar, but has its peak value in the Spearman MSTs.

For Germany the Financials, Industrials and Materials sectors are regarded as important throughout most of the dataset for every MST. However the Materials and Industrials sectors are more central in the Pearson and Spearman MSTs than the

τ. At the end of the time period, the Technology sector has a higher peak in the Pearson MSTs, followed by the Spearman MSTs, but is not particularly important in the τ MSTs. The Consumer Staples, Consumer Discretionary and Healthcare sectors are not particularly central in any of the MSTs, and are relatively similarly expressed.

Compared to the sector degree centrality, the sector betweenness centrality shows a much greater range, and greater disagreement between the MSTs, notably between the rank MSTs. However in general the sectors that are regarded as important in the degree MSTs are also regarded as important in the betweenness MSTs. This could imply that companies tend to be placed in different positions in the different MSTs, even if they have a similar degree centrality

7.3.4 Effects of Gaussian Deviations

In this section we explore the potential reasons for differences between the MSTs, focusing on univariate non-Gaussianity. To measure this we use the Kolmogorov-Smirnov (KS) distance from a Gaussian distribution for each stock for each window and plot this against the absolute normalised node difference, measured by

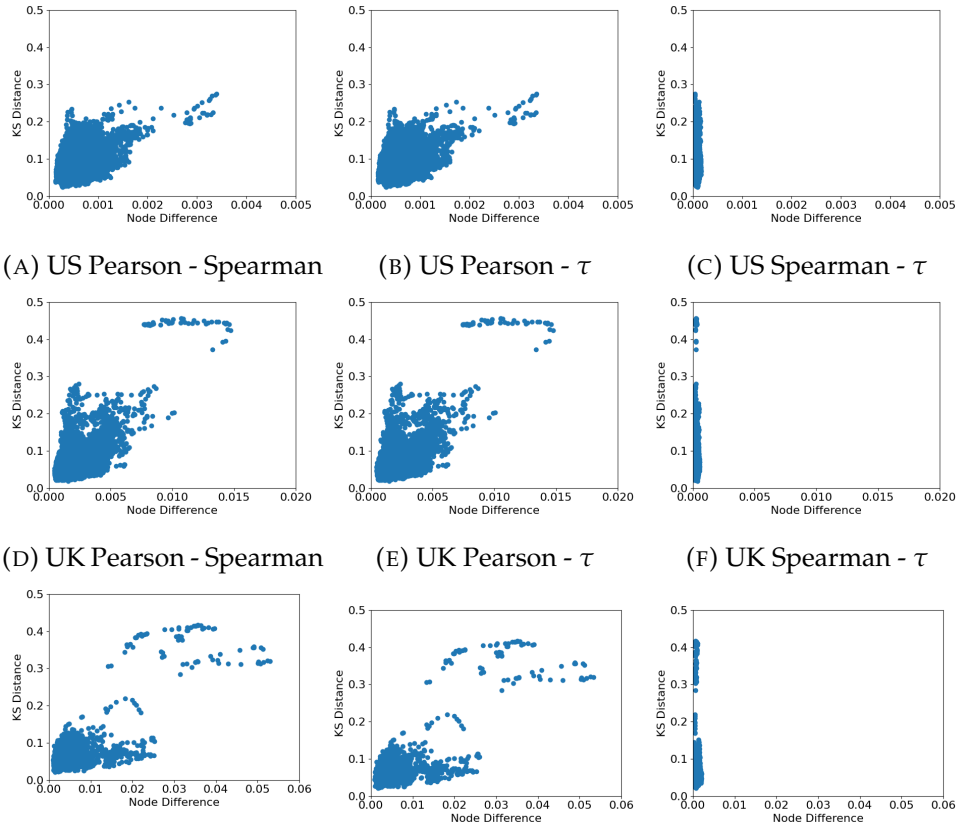
$$\sum_{i=1}^{p} \sum_{j=i+1}^{p} |\frac{C_i^x}{M^x} - \frac{C_i^y}{M^y}| \tag{7.5}$$

where $M^x = \sum_{i=1}^{p} \sum_{j=1}^{p} C_{ij}$ (i.e. the sum of all the correlations in the network), C_i is the ith column of the correlation matrix and C^x and C^y are correlation matrices created from different correlation coefficients. Normalising the correlations ensures that times where the correlations are higher do not distort our measurements. This is done for both the full correlation matrices (Figure 7.8) and the MST filtered correlation matrices (Figure 7.9). To clarify, an MST filtered correlation matrix is the correlation matrix constructed from the MST, where edges in the MST are given the weight of their correlation from the original full correlation matrix and all other correlations are set to zero.

For all countries there is a positive relationship between the KS distance and the distance between nodes in the full correlation matrices when we compare the Pearson and rank networks, indicating that a departure from Gaussianity does increase the difference between the different correlation coefficients. There appears to be no relationship when comparing the rank correlation networks, which is to be expected.

However if we look at the MST figures, the results seem a bit different. For all countries it seems there is relatively little relationship between node difference and KS distance. However since the procedure to create the MST discards the majority of correlations, this could imply that the trees tend to keep correlations that are unaffected by this deviation. Furthermore, since the MSTs only keep large correlations, it could be that the deviations affect smaller correlations more.

To quantify these we show the Spearman correlation between node difference and KS distance in Table 7.1. For all of the full correlation matrices there is some positive correlation between deviation from Gaussianity and the difference between the rank and Pearson correlations. The US and UK also have some very mild negative correlation between the τ and Spearman MSTs. None of the countries have any correlation between the node difference and KS distance with the MSTs. From the scatter plots this is mostly to be expected. The results if we use unweighted MSTs are very similar.



(A) US Pearson - Spearman (B) US Pearson - τ (C) US Spearman - τ

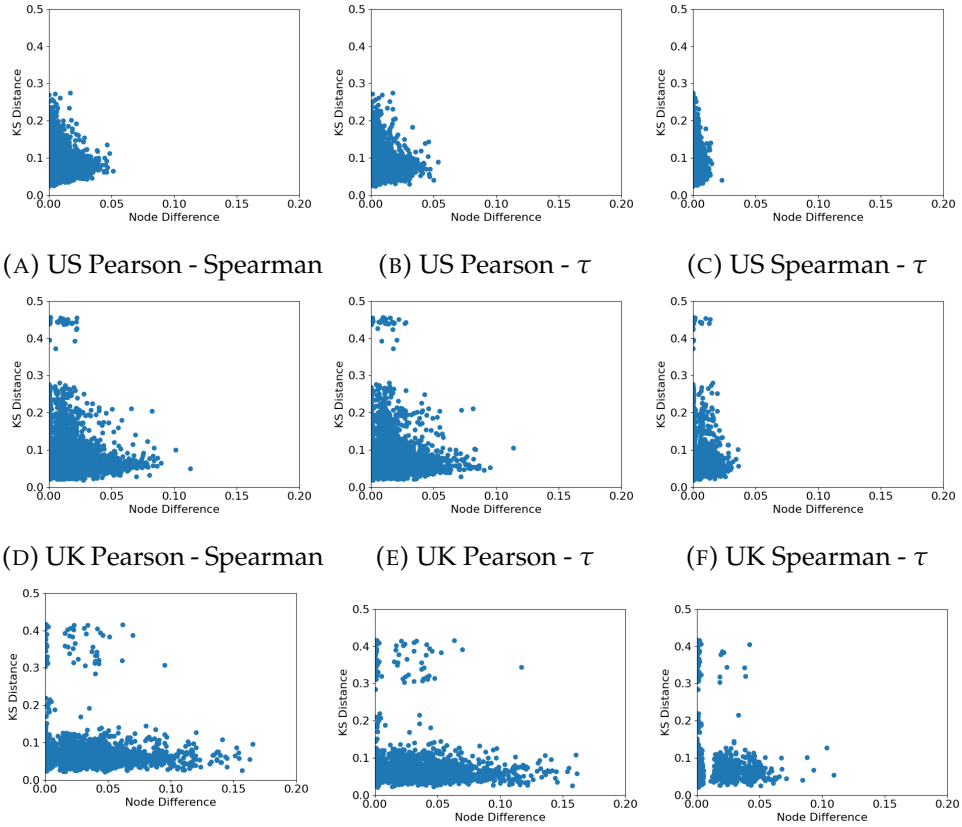
(D) UK Pearson - Spearman (E) UK Pearson - τ (F) UK Spearman - τ

(G) Germany Pearson - Spearman (H) Germany Pearson - τ (I) Germany Spearman - τ

FIGURE 7.8: Scatter plots of the difference for nodes in the full correlation matrix against the KS distance for all 3 countries. There appears to be a positive relationship between node difference and KS distance when comparing the Pearson and rank MSTs, indicating that a deviation from univariate Gaussianity does cause differences

Network	Pearson - Spearman	Pearson - τ	Spearman - τ
US Full	0.221	0.244	-0.114
UK Full	0.409	0.415	-0.104
DE Full	0.331	0.352	-0.075
US MST	-0.005	-0.005	-0.016
UK MST	0.000	-0.008	-0.006
DE MST	-0.013	0.008	0.021

TABLE 7.1: Spearman correlation between node difference and the Kolmogorov-Smirnov distance for the node from a univariate Gaussian. In general a departure from univariate Gaussianity tends to cause differences in the full correlation matrices, but not in the MSTs, potentially due to the filtration procedure.

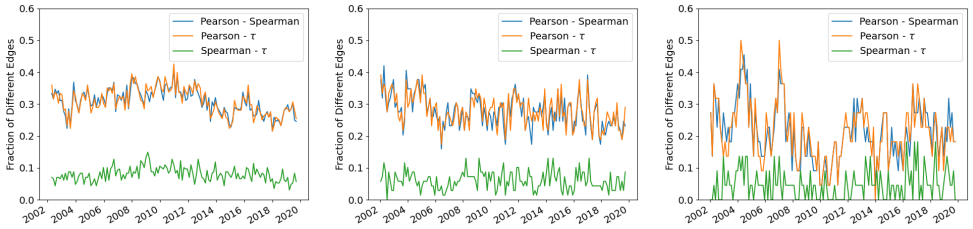


(A) US Pearson - Spearman (B) US Pearson - τ (C) US Spearman - τ

(D) UK Pearson - Spearman (E) UK Pearson - τ (F) UK Spearman - τ

(G) Germany Pearson - Spearman (H) Germany Pearson - τ (I) Germany Spearman - τ

FIGURE 7.9: Scatter plots of the difference for nodes in the MSTs against the KS distance for all 3 countries. There seems to be significantly less of a relationship between the node difference and the KS distance in the MSTs compared to the full correlation matrices. This could be due to the MST procedure discarding the changed relationships.



(A) US (B) UK (C) DE

FIGURE 7.10: Edge difference between the trees over time when quantile normalisation is used to make the asset returns data normal. Comparing this to Figure 7.3 there is a reduction in this difference between the Pearson and rank MSTs, but it is still much higher than the difference between the rank MSTs. This would imply that it is not just a deviation from normality which causes differences between the MSTs.

Our second experiment on this front is to use quantile normalisation to make the distributions of the asset returns normal. We then look at how this changes the differences between the trees. We use 200 quantiles for this, and plot the differences between the MSTs in Figure 7.10. If we compare this to Figure 7.3 we can see the differences between the Pearson and rank methods have been reduced, but that they are still larger than the differences between the rank MSTs.

This therefore implies that, in contrast to the previous results, the departure from univariate Gaussianity does cause differences between the MSTs as well as the full correlation matrices. However it does not explain all of the differences between the MSTs. This would imply that overall there are non-linear relationships present in the dataset that also drive differences between the MSTs, as well as non-normalities.

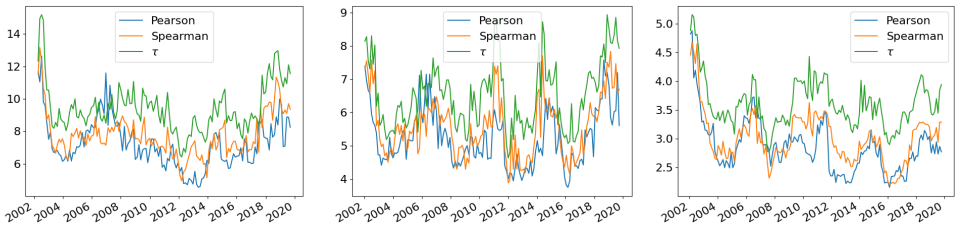
7.3.5 MST Topology

Having studied the stability of the trees over time and the importance of various sectors, we now look if the structure of the MSTs differ using some network measures. We use the leaf fraction (fraction of nodes with only 1 edge), exponent when fitting a power law to the degree distribution, the average shortest path length and mean occupation layer (the mean of all shortest paths from each node to the center of the tree). In this case we take the center of the tree as the node with the largest degree. The plots of these measures over time are shown in Figures 7.11 to 7.13.

From these we can see that irrelevant of the coefficient, the MSTs have similar structure. All of the trees have a heavy tailed degree distribution, with there being a high number of nodes with only one other edge and a small number of edges with a large degree. The structure of the trees does tend to be dependent on market state for all countries. The average shortest length path is slightly longer for the τ MSTs than the Pearson or Spearman, but the τ correlation tends to be slightly smaller for the same value (see Figure 7.1) which would explain the longer paths, as they will have a greater distance.

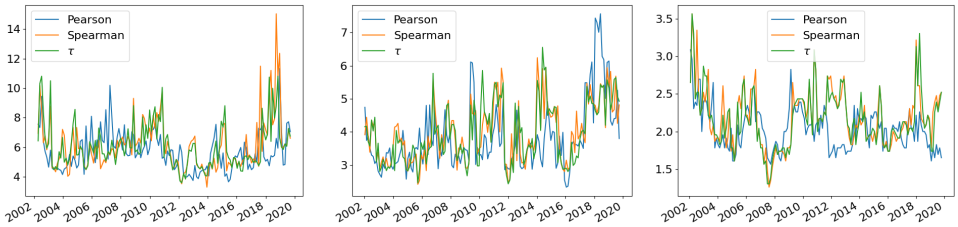
7.3.6 Applications to Portfolio Selection

As mentioned in the introduction, one of the goals in studying the correlations between financial assets is to assess risk. An obvious application is therefore to construct low risk portfolios using said correlations. MST correlation matrices have



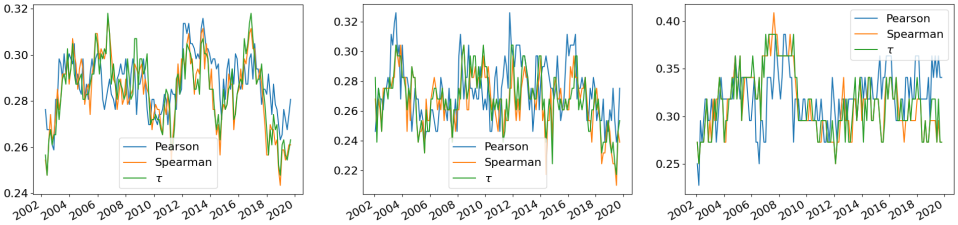
(A) US (B) UK (C) DE

FIGURE 7.11: Average Shortest Path Length



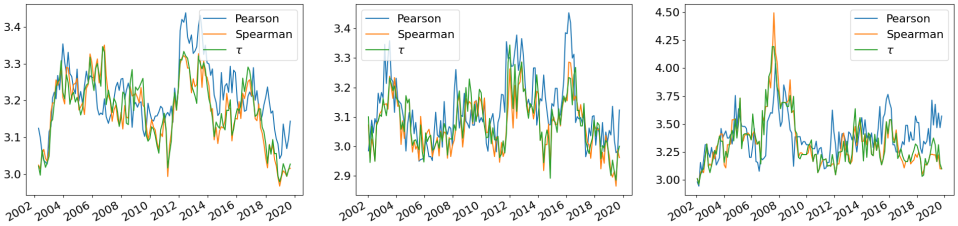
(A) US (B) UK (C) DE

FIGURE 7.12: Mean Occupation Layer



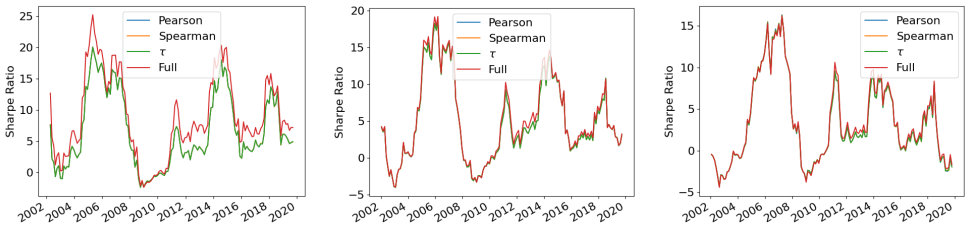
(A) US (B) UK (C) DE

FIGURE 7.13: Leaf Fraction



(A) US (B) UK (C) DE

FIGURE 7.14: Exponent



(A) US (B) UK (C) DE

FIGURE 7.15: Out of sample Sharpe ratio of the portfolios. The full covariance matrix generally has a higher Sharpe ratio than the MST covariance matrices, but the difference is not particularly large. For the larger markets this also comes at the cost of a larger portfolio turnover.

been applied for this in previous work [256], and in this section we look at how the rank MST correlation matrices compare on this front. To do this, we take the MST filtered correlation matrices and turn them into covariance matrices using

$$\Sigma = D^{\frac{1}{2}} C D^{\frac{1}{2}} \tag{7.6}$$

where D is a diagonal matrix containing the variances of the assets.

We have found that sometimes the MST correlation matrices are singular, and therefore add a small amount to the diagonal to ensure the resulting correlation matrix is positive definite as follows

$$\Sigma_* = \alpha\Sigma + (1-\alpha) tr(\Sigma) I \tag{7.7}$$

where $\alpha = 0.9$. This is also applied to the full covariance matrix to assist comparisons. With these resulting covariance matrices, we create minimum risk portfolios by solving the following optimization problem

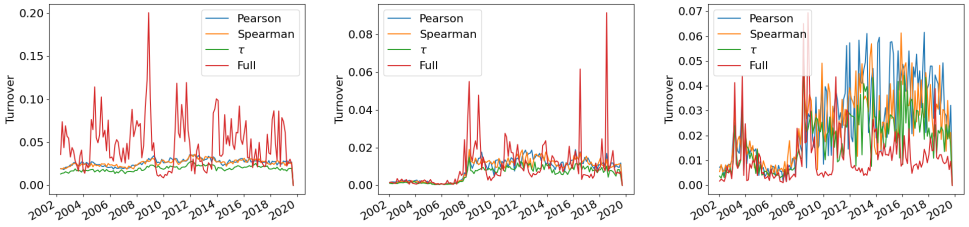
$$\begin{aligned} \underset{w}{\text{minimize}} \quad & \boldsymbol{w}^T \Sigma_* \boldsymbol{w} \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{w} = 1 \\ & w_i > 0 \end{aligned} \tag{7.8}$$

The resulting vector $\boldsymbol{w}$ gives us the weight for each asset.

Firstly we look at the out of sample Sharpe ratios of the resulting portfolios on the following window of data. The results of this are shown in Figure 7.15. The results for all four portfolios are relatively similar and highly affected by market conditions, but in general the full correlation matrices have a higher Sharpe ratio than the MST filtered ones. Next we look at the turnover of the portfolios. Since we have found that the rank MSTs tend to be more stable than the Pearson MSTs, we look at how this translates into improving portfolio stability. We measure this by using the L_1 norm of the difference between two portfolios adjacent in time

$$\sum_{i=1}^{p} |w_{t,i} - w_{t-1,i}| \tag{7.9}$$

This is shown in Figure 7.16. There is a reduction in mean turnover for the MSTs portfolios for the US and the UK, but not for Germany. This could be due to the



(A) US (B) UK (C) DE

FIGURE 7.16: Turnover of the portfolios constructed using the MST correlation matrices. The MST portfolios tend to have a lower turnover than the full covariance portfolios for the US and UK markets, but not for the German market. Out of the three MSTs, the τ portfolios have the lowest turnover.

smaller size of the German markets, causing n to be much larger than p, and therefore the estimation of the full covariance matrix will be much better. For all countries the τ MSTs have the lowest turnover, followed by the Spearman and then the Pearson MSTs. This shows that rank MST covariance matrices may be useful in reducing portfolio turnover when the investor is considering a large number of assets.

7.3.7 MST Robustness

Finally we are interested in comparing the robustness of these correlation coefficients. This can be done using a bootstrap based approach, in a similar manner to Tumminello et al. [263] and Musciotto et al. [177]. Here we create 1000 pseudo-datasets using a circular bootstrap. A circular bootstrap is a type of block bootstrap where we select data from a continuous stretch of time, and if the end of the dataset is reached we wrap round and start back at the beginning. This tends to be more appropriate for time series data compared to the classic bootstrap due to look ahead effects and potential autocorrelation. With these pseudo-datasets we calculate the correlations between assets and construct MSTs from these correlation matrices. Once we have this set of MSTs, we can compare the edges present in them. Ideally if there is no noise, the data is purely stationary and the methods robust all these MSTs would be the same. Of course this is not the case in real life. To run the bootstrap we take the first 1008 days of data and create 1000 bootstrapped datasets of 504 days.

Using these bootstrapped datasets we measure the mean and standard deviation of the difference between the full correlation matrices, the MST filtered correlation matrices (i.e. weighted MSTs) and the fraction of difference in edge presence across the MSTs (i.e. unweighted MSTs). To measure the difference between the full and MST filtered correlation matrices we take a similar approach as to section 7.3.4 and normalise the entries of the correlation matrices to sum to 1 and take the sum of the absolute difference for each entry

$$\sum_{i=1}^{p}\sum_{j=1}^{p}\left|\frac{C_{ij}^{x}}{M^{x}}-\frac{C_{ij}^{y}}{M^{y}}\right| \tag{7.10}$$

where $M^{x}=\sum_{i=1}^{p}\sum_{j=1}^{p}C_{ij}^{x}$. The results are shown in Table 7.2. If we look at the full correlation matrices, the US and Germany both see a reduction in the mean difference when using rank correlation as opposed to Pearson correlation. For the UK there is a reduction in mean difference for the Spearman MSTs, but not for the

Method	MST Weighted		MST Unweighted		Full	
	Mean	S.D	Mean	S.D	Mean	S.D
US						
Pearson	0.835	0.218	0.722	0.056	0.234	0.094
Spearman	0.830	0.209	0.721	0.053	0.175	0.074
τ	0.824	0.210	0.720	0.053	0.174	0.071
UK						
Pearson	0.896	0.214	0.750	0.054	0.296	0.137
Spearman	0.904	0.220	0.749	0.056	0.247	0.123
τ	0.890	0.220	0.747	0.056	0.248	0.123
DE						
Pearson	0.732	0.249	0.690	0.064	0.226	0.101
Spearman	0.700	0.235	0.690	0.063	0.128	0.058
τ	0.665	0.231	0.684	0.062	0.121	0.048

TABLE 7.2: Mean and standard deviation (s.d.) of the difference between full correlation matrices and MSTs constructed from the bootstrapped datasets. For the all of the countries there is a reduction in the mean difference between full correlation matrices when the rank correlation method is used, but there is little reduction for the MST networks, weighted or unweighted.

τ MSTs when compared the the Pearson MSTs. However if we look at the mean difference between the MSTs, the results differ. If we look at the weighted edges there is a slight reduction in difference for the US and Germany, but particularly for the US this is not large. However for the unweighted MSTs there seems to be little to no difference between the MSTs for the US and Germany, and in fact an increase for the UK. From this we would conclude that the MST construction procedure has a larger effect on the robustness of the trees than the correlation coefficient chosen.

7.4 Conclusion

In this chapter we have used the Pearson, Spearman and Kendall's τ correlation coefficients to infer correlation matrices from stock returns from three countries (the US, UK and Germany), constructed minimum spanning trees from these matrices and compared the robustness and evolution of the trees over time.

Looking at the evolution of the trees over the dataset we have found the MSTs constructed using the rank correlations (Spearman and Kendall's τ) change less than Pearson MSTs (notably during times of market stress) and have edges that are maintained for a longer time period over the dataset. The rank MSTs tend to select similar edges, with little variance over time. Compared to this, the difference in edge selection between the rank MSTs and the Pearson MSTs tends to increase during times of market stress. Despite this, the trees tend to have a similar topology, irrelevant of coefficient and this topology tends to vary in a similar way over time for all three methods. Perhaps unsurprisingly, the structure of the rank MSTs is very similar, while they both differ more from the Pearson MSTs.

In general all of the MSTs tend to show broad agreement on which sectors are regarded as important over the entire dataset, but there can be significant disagreements at particular points in time. The rank MSTs show more agreement with each other than either with the Pearson MSTs. The agreement using degree centrality is higher

than using betweenness centrality, indicating that companies tend to be found in different places on the trees in the Pearson MSTs compared to the rank ones.

We then attempt to connect departures from univariate Gaussianity for individual companies to changes in their expression in the MSTs and full correlation matrices. These deviations are correlated with changes in the expression of a company in the full correlation matrices, but not in the MSTs. Furthermore we then use a quantile normalisation method to enforce univariate Gaussianity on each company, and then run the analysis again. From this we find that there is a reduction in difference between the rank and Pearson MSTs, but there is still a much larger difference between them than between the Spearman and τ MSTs, indicating that there could be non-linear relationships involved too.

These MSTs can also be applied for use in portfolio selection. We construct minimum risk portfolios from the MST correlation matrices and compare the resulting portfolios to those produced using the full covariance matrix. The portfolios constructed from the MST correlation matrices tend to have a lower turnover compared to those constructed using the full matrices for the larger markets, but tend to have a slightly lower Sharpe ratio. In particular for the portfolios constructed the MSTs, the τ portfolios have the lowest turnover, while their Sharpe ratio is indistinguishable from those constructed from the Pearson MST correlation matrices.

Finally we use a bootstrap to test the consistency of the correlation matrices inferred and the edges selected in the MSTs. We find that the full correlation matrices constructed using rank correlations mostly have a smaller difference than the full Pearson correlation matrices, but there is relatively little difference in the MSTs, indicating the MST construction procedure has a larger influence on this than the correlation coefficient chosen.

Overall, the MST construction procedure has the greatest influence on the results. Generally the Spearman and Kendall's τ correlation coefficients tend to give similar results, indicating that if computational resources are constrained then calculating the Spearman correlation is sufficient. For financial applications, it could still be worth using rank correlations to quantify asset relationships, as we have found they show improvements with regards to stability. However, many the issues that apply to financial correlation networks independently of the chosen coefficient still apply.

Future work could proceed in several directions. A comparison of mutual information MSTs to these correlation MSTs to see how they differ could be interesting, or exploring different filtration models, for instance the Planar Maximally Filtered Graph. Alternatively these comparisons could be performed with returns data from other countries or assets, perhaps from data that is highly correlated and volatile, for instance for returns from cryptocurrencies or developing nations.

Chapter 8

Stability and Similarity in Financial Networks

8.1 Introduction

In order to reduce risk, many investors wish to own stocks that are unrelated, or even better, negatively correlated. This referred to as a diversified portfolio, and ensures that even if a certain portion of their portfolio declines, the overall value remains the same. Studying how correlations vary between assets allows us to understand how to approach the construction of diversified portfolios. Therefore in this chapter we study how the stability of correlation networks and the similarity between nodes in said correlation networks varies during different market states.

The similarity of nodes in the network is important due to the concept of creating diversified portfolios. If the nodes in a correlation network are very similar, this implies that it is not possible to construct a diversified portfolio from said nodes. Furthermore if the similarity varies over time, assets that were once dissimilar can become similar, changing from a good choice for a diversified portfolio to a poor one.

The stability of the networks can also be related to portfolio construction. If the networks are unstable, this makes it challenging to predict the future performance of assets, and by extension, portfolios constructed from said assets. Therefore understanding the stability of the networks is also important for exploring the possibility of constructing portfolios.

8.2 Methods

8.2.1 Network Construction

To construct this network, we calculate the correlation matrix of the log returns of the stocks. This correlation matrix is then used as the adjacency matrix for the correlation network, with the assets becoming nodes and the correlation between two assets being the weight on the edge between them, in the same manner as in the rest of this thesis.

Correlation networks by default are complete and weighted. Furthermore, these weights can be negative. Both of these can cause problems when using some similarity measures, which tend to either implicitly or explicitly assume that the networks are sparse and only contain edges with positive weight. This is specifically the case for heterogeneity and Katz similarity. We therefore use a filtration method to solve these issues. A description of the various methods available can be found in Chapter 2. For our analysis we choose the Planar Maximally Filtered Graph (PMFG) [260] to filter the full networks, and then apply the similarity measures to the filtered

networks. We also take the absolute values of the correlations as input into the PMFG procedure as the similarity method can fail to converge when negative edges are present.

8.2.2 Network Stability

To measure the stability of the correlation networks we use the L_2 difference between the eigenvectors that correspond to the largest eigenvalue (the leading eigenvectors) from networks adjacent in time. These are also normalized so the entries add to one. This is defined as:

$$||\frac{\boldsymbol{v_t}}{\sum_{i=1}^{p} v_{t,i}} - \frac{\boldsymbol{v_{t-1}}}{\sum_{i=1}^{p} v_{t-1,i}}||_2 \tag{8.1}$$

where p is the number of companies in the network, $\boldsymbol{v_t}$ is the leading eigenvector at time t and $||\boldsymbol{x}||_2$ is the L_2 norm of vector $\boldsymbol{x}$.

The entry that corresponds to a node in this eigenvector can be interpreted as the centrality of said node in the network. This difference therefore measures how much the centrality of each node changes from one network to the next - so effectively we measure if corresponding networks agree on which nodes are regarded as important.

8.2.3 Network Similarity

Many methods have been proposed to measure node similarity in networks. Broadly these fall into three categories, structural equivalence, regular equivalence or automorphic equivalence [185].

Structural equivalence is perhaps the easiest to understand and calculate, we simply measure how many neighbours two nodes have in common [155]. This can be quantified by measuring the similarity between the appropriate rows of the adjacency matrix with a vector distance measure e.g. Pearson correlation. To quantify the amount of structural equivalence in our networks, we use two measures. The first is the mean L_2 distance between the rows of the correlation matrix, defined as:

$$\frac{1}{p(p-1)} \sum_i \sum_{j \neq i} ||\boldsymbol{C_i} - \boldsymbol{C_j}||_2 \tag{8.2}$$

where $\boldsymbol{C_i}$ corresponds to column i of the correlation matrix.

The second is the mean cosine distance, defined as:

$$\frac{1}{p(p-1)} \sum_i \sum_{j \neq i} 1 - \frac{\sum_{k=1}^{p} C_{ik} C_{jk}}{||\boldsymbol{C_i}||_2 ||\boldsymbol{C_j}||_2} \tag{8.3}$$

We choose two measures due to the varying amount of correlation present at different times. The cosine distance is normalized while the L_2 distance is not. This allows us to see if this normalization affects the results.

We may also be interested in methods that measure nodes that hold similar positions in networks, despite not sharing any neighbours. This can be measured by *Regular equivalence*, defined by [281]. Here, nodes are similar if they are connected to other nodes that are themselves similar. To measure this we use a form of Katz similarity, defined as

$$(D - \alpha A)^{-1} D \tag{8.4}$$

where D is the degree matrix (a matrix with the degree of each node on the diagonal), A the adjacency matrix of the network, and α a 'dampening factor'. The Katz similarity

allows paths of all lengths to contribute to the similarity. α controls how longer paths contribute, a small α (i.e. close to 0) means longer paths will contribute little, while an α close to 1 ensures all paths will be given the same weight. This particular form of Katz similarity is normalized so that nodes of a higher degree are not biased to be more similar. We choose α to be 0.05.

Automorphic equivalence is formally defined as "two nodes of a graph are automorphically equivalent if all the nodes can be re-labeled to form an isomorphic graph with the labels of u and v interchanged" [99]. Calculating this can be computationally expensive, as it requires computing graph automorphisms. Furthermore, it is difficult to define for correlation networks as it does not depend on the weight of the edge. Due to these issues, we do not use automorphic equivalence.

If the reader is interested in more detail, we direct them to the following reviews [156] [197], both of which contain sections reviewing node similarity measures.

An alternative approach to measuring the similarity in the network is to look at the distribution of node centrality. This again assume that nodes that share a similar structure are more similar. Generally, these are based on the degree distribution. The first piece of work on this topic is by Bell [18], who proposed to use the variance of node degrees as a measure. Intuitively, if this is large then the nodes have a large difference in degree, and therefore would be expected to be dissimilar. If it is small then the nodes all have similar degrees, and therefore are more similar. We use this approach to measure the similarity in the full correlation networks. As previously mentioned, the entries in the leading eigenvector correspond to the eigenvector centrality of the nodes, and so we measure the standard deviation in the entries of the normalized leading eigenvector.

We also use a formulation proposed by Estrada et al [77]. This defines heterogeneity as follows:

$$H = \frac{\sum_{i,j \in E} (k_i^{-1/2} - k_j^{-1/2})}{p - 2\sqrt{p-1}} \tag{8.5}$$

where k_i is the degree of node i and E is the edge set of the graph. If the network is completely uniform, with all nodes having the same degree, the heterogeneity will be 0. On the other hand if there is a great amount of variation in node degree (for instance a star graph), the heterogeneity will be close to 1. This cannot be used on the full correlation networks as it assumes the networks are unweighted, and so we only apply this method to the PMFG networks.

8.2.4 Community Detection

We can also study the similarities and differences between nodes in a network using community detection. To do this we maximize the modularity of the network using the Louvain algorithm [25]. The modularity for a network with adjacency matrix A and a vector of community assignments $\boldsymbol{c}$ is [184]

$$Q = \frac{1}{m} \sum_i \sum_j (A_{ij} - \frac{k_i k_j}{m}) \delta(c_i, c_j) \tag{8.6}$$

where $m = \sum_i \sum_j A_{ij}$ and $\delta(c_i, c_j)$ is the Kronecker delta function, equalling 1 when $c_i = c_j$ (i.e. nodes i and j are in the same community and 0 otherwise).

The modularity is maximized in a greedy manner. Firstly each node is assigned to its own community. Next we iterate through the nodes, calculating the loss in modularity of removing it from said community, and the gain for adding it to a

different community, as follows:

$$\delta Q = \frac{\sum_{\text{in}} + 2k_i}{2m} - (\frac{\sum_{\text{tot}} + k_i}{2m})^2 - (\frac{\sum_{\text{in}}}{2m} - (\frac{\sum_{\text{tot}}}{2m})^2 - \frac{k_i}{2m}^2) \tag{8.7}$$

where $\sum_{\text{in}}$ is the sum of weights of all the edges inside the community node i is being moved into, $\sum_{\text{tot}}$ is the sum of weights of the edges to the community. The node is assigned to the community that gives the largest gain in modularity. This process is repeated until the overall gain in modularity is no longer positive.

This formulation of modularity does not permit negative edges, and so we simply add one to every edge in the network. While there are more elegant methods (e.g. [93] [157]), we find this method gives comparable results with a simpler implementation.

8.3 Software and Data

The data we use is downloaded from Yahoo Finance. For the UK data we use the FTSE100 companies, for the US returns we use the S&P500 companies, and for the German data we use the DAX30 companies. We use returns from 2000/03/01 to 2019/10/21. For each dataset, any company missing more than 10% of its data is removed, and any missing values are filled forwards from the first good value. If the values are missing from the start we backfill from the first good value. We use a window of 252 trading days (1 trading year) for Germany and the UK, and 504 (2 trading years) for the US due to the larger sample size and slide it along 30 days at a time. This results in 5065 days of return data for 70 companies for the UK, 5068 days for 23 companies for Germany, and 4790 days of return data for 229 companies for the US. Each company is tagged with a sector from the GICS classification using information from Bloomberg. This places each company into 1 of 11 sectors, Information Technology, Real Estate, Materials, Telecommunication Services, Energy, Financials, Utilities, Industrials, Consumer Discretionary, Healthcare or Consumer Staples.

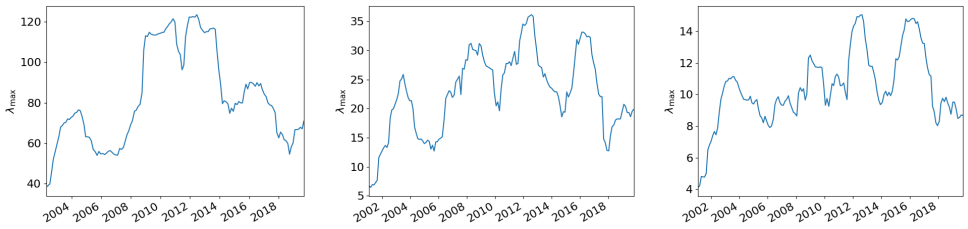
We make use of Python, NumPy and SciPy [193] for general scripting, pandas [118] for handing the data, statsmodels [227] for some of the statistical analysis, matplotlib [110] for plotting, TopCorr (`https://github.com/shazzzm/topcorr`) for construction of the PMFGs, Cytoscape [229] for visualization of the networks and Networkx [96] for the network analysis.

8.4 Results

8.4.1 Periods of Disruption

Firstly we look at the largest eigenvalue of the correlation matrices to demonstrate when the networks detect a period of market disruption. The largest eigenvalue measures the intensity of correlation in that network and should be larger during periods of market disruption. This is shown in Figure 8.1.

The US market has a far larger eigenvalue than either the German or UK markets, but it is a far larger matrix. It also has more obvious peaks in intensity - notably in 2003, 2009, and 2012. These relate to the dot com crash, financial crash, and European sovereign debt crisis respectively [79]. The UK has similar peaks in 2003, 2008, 2012 and 2016 which contains periods of disruption during the financial crisis and the Brexit vote of 2016. Germany is noisier, but with noticeable peaks in 2003, 2012, and 2016. The two largest peaks are in 2012 and 2016, which could relate to the



(A) US (B) UK (C) DE

FIGURE 8.1: Largest eigenvalue of the correlation matrices inferred from the market returns. 2009 and 2012 are noticeable peaks in the US markets probably due to the financial crisis from 2007 onwards, while the dot com crash of 2003 is also visible. The other markets seem to be noisier, although the UK has peaks in 2003, 2008, 2012 and 2016 - the latter of which could be related to the Brexit vote. Germany has peaks during 2003, 2012 and 2016 too, although these are smaller (note scale on axis) but this could be due to the smaller market

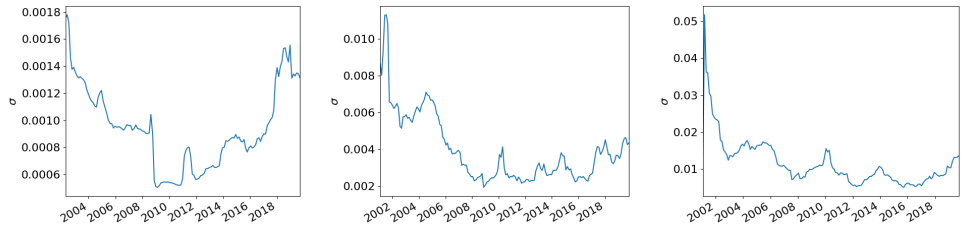
(A) US (B) UK (C) DE

FIGURE 8.2: L_2 difference between the eigenvector centrality of adjacent networks of the US market. Here our goal is to measure the stability of the underlying structure of the correlation matrices over time. Particularly for the US it is clear than the differences drop during 2009 and 2012, which are both times of market stress. With the UK we can see that there are minimums in 2009 and in 2016, which again are times of market stress for the UK, while Germany has low points during 2008, 2012 and 2016.

Eurozone sovereign debt crisis and the Brexit vote, both of which have deep ties into the German economy [46].

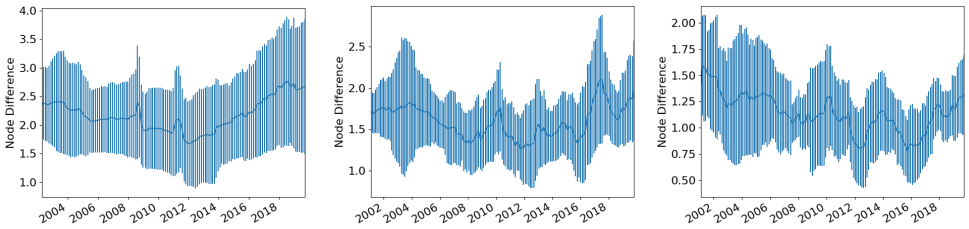
8.4.2 Stability

Next, we look at the difference between the leading eigenvectors of the networks. The results of this are shown in Figure 8.2. For the US, we see there is an obvious dip from 2009 to 2011, showing that surprisingly the system seems more stable during this time of disruption. If we look at values from times of growth and stability (e.g. 2006) the difference between adjacent windows is around 0.45, indicating most of the leading eigenvector changes during these times. Before large drops (notably in 2009 and in 2011), there are small peaks in the difference. The UK has a slightly different story, with the start of the dataset showing a large change, following by the difference decreasing until 2008. From this year it is much more stable. There are small peaks in 2004, 2010, 2015 and 2017. It is interesting to note how the stability increases after 2004 until 2008, levelling off at around 0.3. Germany is again distinctive. At the start of the dataset there are large changes in the entries of the



(A) US (B) UK (C) DE

FIGURE 8.3: Standard deviation of the normalized leading eigenvector for the networks over time. This is a measure of the similarity in the networks - if the standard deviation is low then the nodes all have a more similar centrality and therefore should be more similar. We again note the drops in 2009 and 2012 for the US market, and in 2012 and 2016 for the UK markets.



(A) US (B) UK (C) DE

FIGURE 8.4: Mean L_2 difference between the nodes in the network, defined in equation 8.2. Error bars are the standard deviation. For the US markets the financial crisis is particularly noteworthy with there being a significant drop in difference between 2009 and 2010, but peaks in 2008 and 2011. With the UK there is a peak of difference during 2017 when there was a strong bull market. Germany has minimums during 2008, 2012, and 2016.

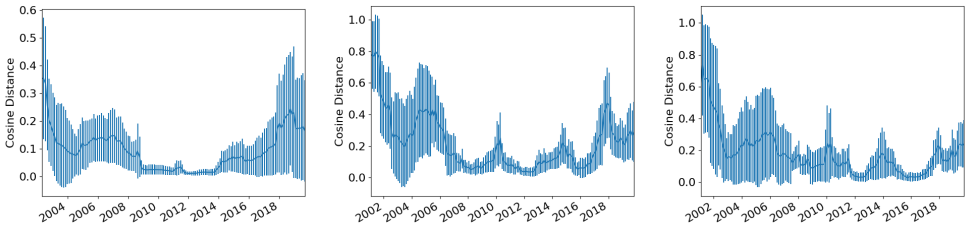
leading eigenvector, but it becomes far more stable as time goes on, with minimums in 2008, 2012, and from 2015 to 2017.

8.4.3 Full Network Similarity

In this section, we study the similarity present in the full correlation networks. The first measure we use is the standard deviation of the entries of the normalized leading eigenvector - which as previously mentioned relates to the eigenvector centrality of a node in the network. If this value is large then there is a big spread in the importance of the nodes and therefore they could be dissimilar. If it is small then they all have similar centralities and therefore could be similar. A plot of this over time is shown in Figure 8.3.

Starting with the US, we note that the standard deviation is very high in 2002 and 2018. Again we see this drop from 2009 - 2010, during the financial crisis. For the UK, we see drops during 2009, 2012, and 2016, which as previously mentioned correspond to large negative macroeconomic effects. Germany has a maximum in 2002 like the others, and minimums during 2008, 2012, and 2016, but the movement seems to be more subtle.

Next, we look at how the structural similarity of the networks varies over time. To start with we look at the L_2 distance between nodes. A plot of this over time is



(A) US (B) UK (C) DE

FIGURE 8.5: Mean cosine difference between the nodes in the network. Error bars are the standard deviation. We see an even greater change in cosine distance between nodes during the financial crisis than L_2 distance, particularly for the US.

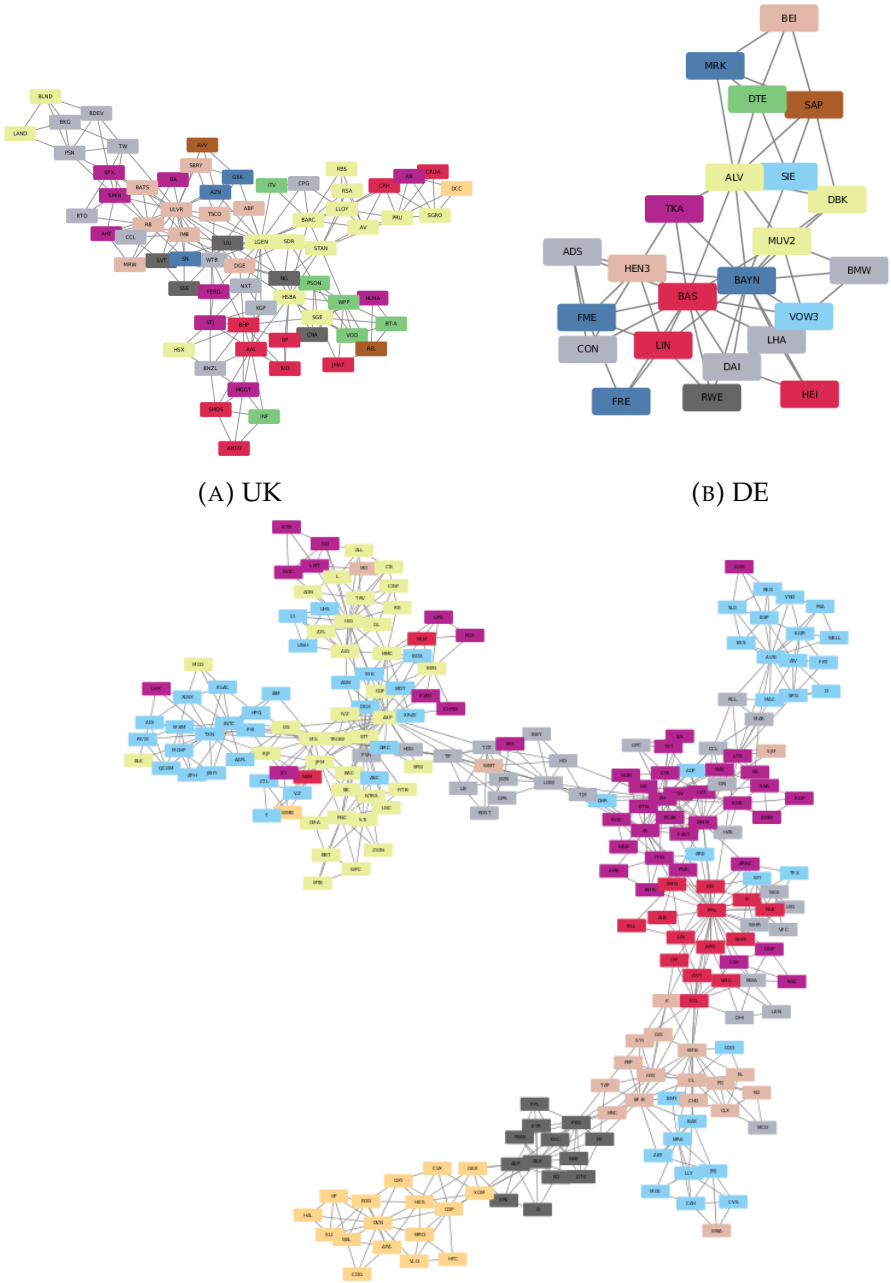
shown in Figure 8.4. For the US market, we can see the difference drops from 2008 to 2009 and from 2010 to 2012, which is consistent with our other networks measures of the perception of a crisis - implying the nodes do become more similar during times of disruption. With the UK market, we can see a small drop during 2012 and 2016, which as mentioned are times of difficulty. Interestingly, there is a large rise during 2017. This time was a period of growth for the FTSE100 due to the weakened pound making UK company currency holdings worth more. This does line up with our theory that bull markets correspond to times of greater difference between nodes, but since this is a general large macroeconomic effect it is worth mentioning that the gains do not seem to be as evenly spread out. Finally, we look at the results for the German market. Again the differences here are smaller than for the UK and US markets, but we do notice a drop during 2012 and 2016.

We then move to our other measure of structural similarity, the cosine distance. Plots of how this varies over time are shown in Figure 8.5. The effects here are even starker than for the L_2 distance, with both the mean distance and the standard deviation dropping significantly during the financial crisis. This larger drop is due to the effects of normalization. Since correlations are larger during times of stress, normalization reduces the relative distance between them, even if the absolute difference is larger.

8.4.4 PMFG Analysis

In this section, we study the properties of the PMFGs constructed from the stock returns of the three countries. Firstly we show example networks from the first window of data for all three countries. These are shown in Figure 8.6.

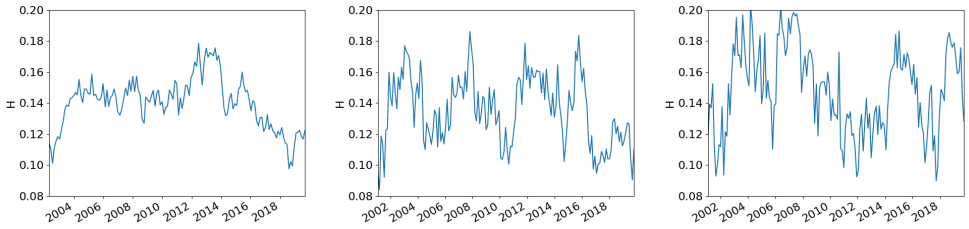
Next, we look at the degree heterogeneity. This is shown in Figure 8.7. From this we first note that the PMFGs are not particularly heterogeneous, indicating a relatively uniform degree distribution in general. This could be due to the construction procedure, which does enforce a certain structure on the network. For the US the heterogeneity is stable, being around 0.14 for most of the dataset. There is an increase from 2012 - 2014 and minimums in 2003 and 2019. This is not particularly easy to connect to any market crashes. For the UK there is significantly more variation. There are peaks in 2003, 2007, 2011, and 2016 and a noticeable drop from 2009 - 2011, which could be connected to the financial crisis. Germany is noisier still, with a great deal of variation in the heterogeneity. The increase in variation as we move from the US to the UK to Germany could be due to the decreasing size of the markets, making smaller changes more visible.



(A) UK (B) DE

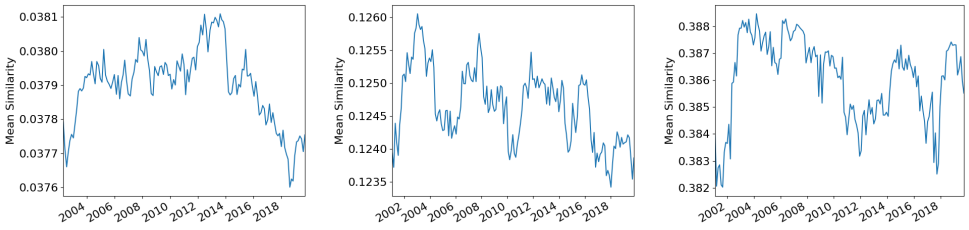
(C) US

FIGURE 8.6: Example PMFGs constructed from the first window of stock returns for all three countries. Nodes are coloured according to sector membership.



(A) US (B) UK (C) DE

FIGURE 8.7: Degree Heterogeneity of the PMFGs over time (this is defined in equation 8.5). For the US this is relatively stable in time, making it difficult to connect to a market state. The UK and Germany show much more variation.



(A) US (B) UK (C) DE

FIGURE 8.8: Mean Katz Similarity in the PMFGs over time. This shows a very similar trend to the heterogeneity (see Figure 8.7), indicating the two are picking up on similar trends

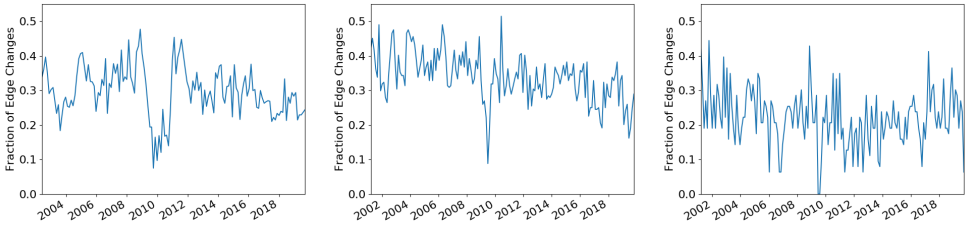
With this in mind, we next study Katz similarity in the PMFGs. A plot of its mean value over time is shown in Figure 8.8. This shows a very similar trend to the heterogeneity for all of the countries, indicating the two are picking up similar trends. Again this makes it difficult to connect to specific times for the US and Germany, but the UK can be linked to market states.

In terms of which nodes are regarded as similar to which, we found that if a node had a high similarity to another, they tended to be in the same sector, and in general the mean similarity was higher between nodes in the same sector. However, the full sector structure is not immediately visible. To quantify this we look at the point biserial correlation between the similarity between two nodes and whether they are in the same sector. Across the full dataset, this is 0.257 for the US, 0.245 for the UK, and 0.108 for Germany, indicating that the sector structure is not particularly strong. In fact, if we do the same for the full correlation network, we get 0.417 for the US, 0.393 for the UK, and 0.484 for Germany. Therefore the sector structure is stronger in the correlation matrices than the similarity matrices.

Finally, we look at how the stability of the PMFGs varies over time. To measure this we use the number of edge changes between networks adjacent in time. The results of this are shown in Figure 8.9. All three graphs show a sharp drop in edge changes in 2009, and for the US this low is maintained until 2011. The US also shows a drop in 2003, but for the other two countries there is no discernible pattern.

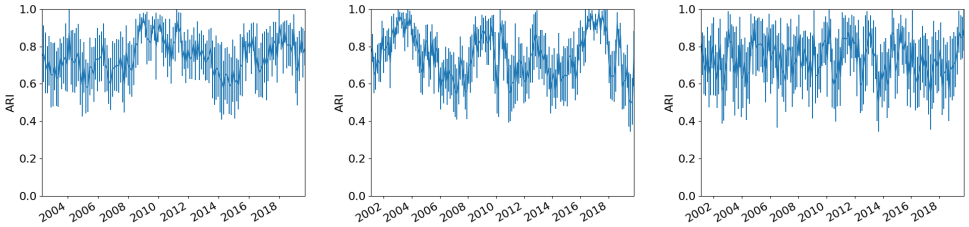
8.4.5 Community Detection

In this section, we use the Louvain modularity maximization algorithm on the full correlation networks to perform community detection. We are interested in four



(A) US (B) UK (C) DE

FIGURE 8.9: Number of edge changes between PMFGs adjacent in time. All three show a sharp drop during 2009, indicating an increase in stability during this time, however beyond this there does not seem to be any other patterns.



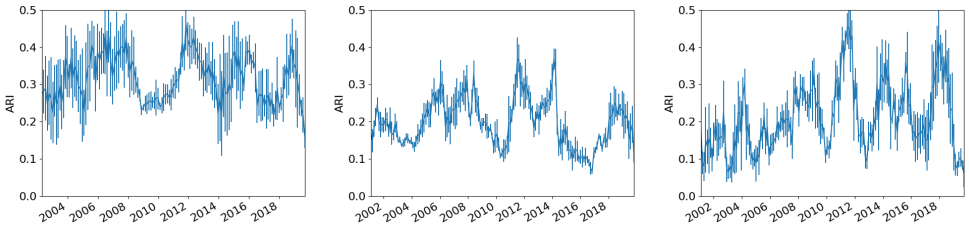
(A) US (B) UK (C) DE

FIGURE 8.10: Mean Adjusted Rand Index from the 10 runs on the same network. Since the algorithm is greedy a different result will be reached each time, but if there is structure to be discovered we would expect similarities between the different runs. In general the rand score is high and therefore the communities produced are quite similar. It does also vary over time, with periods of market stress seeming to have higher consistencies (see 2008-2010 in US and UK, 2016 in the UK)

measures, the number of communities in the network, the adjusted rand score between the 10 runs on each network (referred to as the community consistency), the adjusted rand score when comparing the community assignments to the underlying sector assignments (using the GICS assignments) and the adjusted rand score when comparing the assignments from one network to the next in time - referred to as the community stability over time.

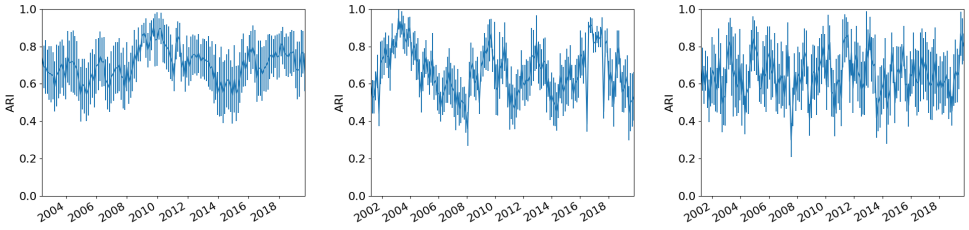
Firstly, we look at how consistent the community detection is when run on the same network. This is shown in Figure 8.10. The communities produced seem to be consistent, with the US mean being at 0.751, UK mean at 0.732 and German mean at 0.730. This consistency is also time-varying - for both the US and the UK there are peaks during the financial crisis and the UK also has other peaks in 2004 and 2016. This would indicate that there is more signal during these times for the community detection procedure to uncover.

The adjusted rand score of the networks over time is shown in Figure 8.11. Interestingly, the level of success in uncovering this underlying sector structure is relatively small, showing that perhaps macroeconomic effects are more important than sector relationships. For the US markets, both the mean and standard deviation of the rand score decrease during 2003 and from 2008 - 2010, indicating that the market distress causes companies to behave more similarly irrelevant of whether they are in the same sector. The UK shows a similar trend, with a decrease in the rand index from 2008 - 2010. There are



(A) US (B) UK (C) DE

FIGURE 8.11: Adjusted Rand Index of the community assignments when compared to the known GICS underlying sector assignments. In general there is not a great level of success in uncovering this underlying sector structure, perhaps indicating macroeconomic effects are more important than sector specific effects. There does also seem to be dips during times of market stress, with both the mean and standard deviation decreasing during these times in the US and UK markets.



(A) US (B) UK (C) DE

FIGURE 8.12: Stability of the communities over time (measured as the adjusted rand index of the networks from one period to the next). For all the markets this is a noisy measure, but the communities do seem to be relatively consistent. In particular it increases for the US during 2003 and from 2008 - 2010, which fits in with our hypothesis that the markets are actually more similar during times of disruption

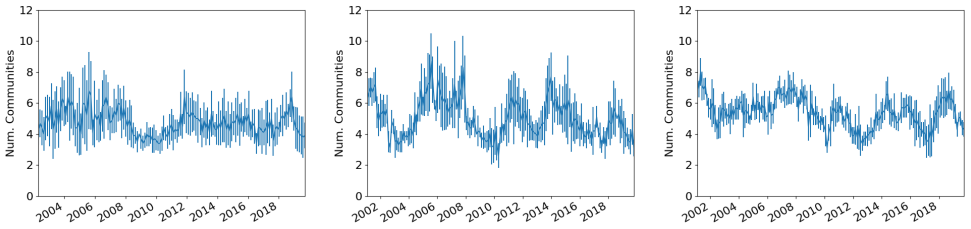
also lows in 2004 and 2017. Germany shows a great deal of variation compared to the other two countries, with maximums in 2012 and 2018, and a large number of lows. It is difficult to connect this to any market conditions.

The community stability over time is shown in Figure 8.12. We see peaks for the US markets from 2008 - 2010 indicating that surprisingly, there is more signal during times of disruption than calmer times. The UK has a similar trend, showing an increase in community stability over the financial crisis, but this is not quite as clear as it is for the US. There is generally more variation compared to the US. The German stability is very noisy, with no discernible time varying structure.

The number of communities in each network is shown in Figure 8.13. If the markets are becoming more similar we would expect fewer communities, so we should see a drop during times of disruption. This does seem to be the case, with drops for the US and UK between 2008 and 2010, and for the UK and Germany during 2012 and 2016.

8.4.6 Correlation Between Network Measures and Volatility

Finally, we look at the Spearman correlation between the various market measures and the volatility of the index. We also conduct a 2 sided hypothesis test to ensure



(A) US (B) UK (C) DE

FIGURE 8.13: Number of communities detected for each of the markets over the dataset. There does seem to be fewer communities during times of market stress, with perceptible lows for the US and UK between 2008 and 2010, and for the UK during 2016. Germany has drops during 2002, 2012 and 2016.

Network Measure	Correlation with Market Volatility		
Country	US	UK	DE
Largest Eigenvalue	0.593	0.494	0.483
Leading Eigenvector Difference	-0.404	-0.304	*0.067*
L_2 Distance	-0.311	*-0.122*	*-0.059*
Cosine Distance	-0.500	-0.289	0.202
Leading Eigenvector Standard Deviation	-0.411	-0.291	*0.069*
Adjusted Rand Index	-0.444	-0.230	-0.228
Community Stability over Time	0.400	0.416	*-0.097*
Number of Communities	-0.502	-0.513	-0.299
Community Consistency	0.378	0.430	-0.270
Heterogeneity	*0.103*	0.335	*-0.152*
Mean Katz Similarity	*0.111*	0.583	*-0.148*
PMFG Edge Changes	*-0.139*	*-0.024*	*-0.030*

TABLE 8.1: Spearman correlation of the various network measures with the volatility of the index. The US and UK markets clearly have strong relationships between the volatility of the index and the network measures, while for Germany this is not the case.

the correlation is significant. To measure the volatility we use the standard deviation of the log returns of the index over the time window the network is inferred from. The results for all three countries are shown in Table 8.1. Non-significant correlations ($p > 0.05$) are in italics.

Since the UK and US have similar results we focus on them first. Both have a strong correlation between the volatility of the index and the largest eigenvalue, showing this is a reasonable measure of the disruption present in the market. Both also show a negative correlation between the L_2 distance, cosine distance, eigenvector centrality difference, and the market volatility indicating that the nodes do become more similar during times of disruption. The leading eigenvector also seems to be more stable during these times of disruption. However, it is important to state the correlation between the L_2 distance and volatility is only significant for the US market. The German picture is different, having only two significant correlations between the volatility and the network measures. In this case, this is the largest eigenvalue and the cosine distance, both of which share a positive correlation with the market volatility.

For the community detection, we have a negative correlation for all markets

between volatility and adjusted rand index, which is something that fits in with our hypothesis - if the nodes become more similar then the underlying sector structure matters less. The US and the UK have positive correlation between volatility and community stability and consistency, and all markets have negative correlation between the number of communities and the volatility.

With the PMFG measures (heterogeneity, Mean Katz Similarity, and the edge changes) we see that only the UK has any significant correlations, with heterogeneity and Mean Katz Similarity having positive correlations with the market volatility. It is somewhat surprising that only the UK shows any relationships between volatility and network measures from the PMFGs. This could be due to the filtration procedure, as it discards the majority of the correlations present, and it could be some of these smaller correlations that are causing these effects in the full networks.

Throughout this investigation, it appears the US and UK markets show relatively similar effects, while Germany deviates significantly. Part of this could simply be due to the significantly smaller size of the German market (23 companies vs 70 and 229) but it is also known the German economy has a significantly different set up the US and the UK. Both the UK and the US have much larger financial sectors than Germany, which places greater emphasis on their stock markets.

8.5 Discussion and Conclusion

In this chapter, we have inferred correlation networks using returns data from the US, UK, and German markets, and investigated how the stability and similarity between companies changes over time. To this end, we have used both full correlation networks and those filtered using the PMFG. We have also performed community detection on the full correlation networks.

Firstly we study the full correlation networks. To measure stability we use the L_2 norm of the difference between leading eigenvectors and the stability in the communities detected from adjacent networks. We find that the leading eigenvectors tend to change less during times of market disruption for the UK and US markets, and that the community consistency and stability increases. This is perhaps surprising, as we would expect the networks to change more during these times. Next, we explore how the similarity between nodes changes, measuring this using the standard deviation of the leading eigenvector, the L_2 and cosine distances between the rows of the correlation matrix, the number of communities in the community detection, and the success of recovering the underlying sector structure. We find that all these measures drop during times of distress for the US and UK markets indicating that the nodes do become more similar during times of market disruption, but that for the German market the only measures with significant relationships with the market volatility are the cosine similarity and the rand index.

For the PMFGs the results are different. Using the mean Katz similarity and the heterogeneity to measure similarity in these filtered graphs, we firstly find that these measures show very similar trends. For Germany and the US, these measures are difficult to link to particular market states, and this is reflected in their low correlation with the index volatility. This is not the case with the UK, which shows a strong correlation with market volatility. Measuring the fraction of edge changes, all three show a sharp drop in the number of edge changes in 2009, but beyond that, there is little relation with market volatility. This is perhaps surprising, although this could be due to the PMFG construction procedure discarding many of the correlations

that are changing. While nodes are more likely to be considered similar to those in the same sector, the relationship is not strong.

For financial practitioners, our results on node similarity demonstrate some of the difficulties of portfolio diversification. Even when an investor creates a portfolio of supposedly dissimilar, diversified stocks during a time of market calm, this actually changes during a time of market stress, with the stocks becoming correlated and similar, and therefore no longer diversified. However, once the crisis period is over, the stocks returns to appearing dissimilar and diversified. Our results on network stability are more difficult to interpret from this perspective. Some authors have noted that there are significant structure changes during a financial crisis [283] [147], but this is not obvious in these networks. The increased stability could imply that prices may be more predictable, but if the end of the crisis cannot be easily predicted, this may lead to a model which fails suddenly. We suggest that future research could investigate this further, ideally with input from financial experts.

Chapter 9

Conclusion and Future Work

9.1 Summary

In this thesis we have approached the subject of empirical finance from a network science perspective. Underlying the studies is the goal of capturing dependencies between asset returns using data which is both time varying and noisy. Robust methods are thus required. We explore partial correlations and various related methods of inference, how they impact portfolio selection, turnover and how the inferred networks vary in times of market stress. Furthermore we also explore the use of rank correlation methods to see if they can add robustness to the inferred networks. We briefly summarize the main findings of the work.

Firstly, in Chapter 2 we provide a review of the literature on financial network inference and analysis, including the different methods used to quantify asset similarity (e.g. correlation, mutual information), methods to filter this information (e.g. topological, random matrix theory), applications of these networks (e.g. portfolio selection, economic sector influence) and clustering methods.

In Chapter 3 we provide a review of the various precision matrix estimation methods, and a Python implementation of a select number. We then investigate how successful a variety of methods are at uncovering a precision matrix from data. Surprisingly we find that simple thresholding of covariance matrices is competitive with more sophisticated formulations of precision matrix inference.

We then apply a variety of estimators to financial returns to obtain precision matrices for portfolio optimization in Chapter 4. We find that the sparse precision matrix estimators can improve portfolio turnover, and the sparsity constraint on the precision matrix can be interpreted as a diversification constraint on the portfolio. However they do not tend to improve out of sample risk when considering minimum risk portfolios.

Pearson correlation is the most ubiquitous method to infer edge weights between assets, but does create the issue of indirect correlations - if A and B and A and C are correlated, this could cause B and C to be correlated. Therefore in Chapter 5 we explore the use of partial correlation to remove such indirect relationships. Creating a partial correlation network of the S&P500 we find that the networks produced tend to be more unstable than their correlation based counterparts, and generally have a less stable community structure with a larger number of communities. This is further verified in appendix A. Stocks from the peripheries of both correlation and partial correlation networks tend to be less risky (there is mild negative correlation with the out-of-sample risk).

Ideally we desire a network to be sparse, as these networks are easier to interpret and analyze. Therefore in Chapter 6 we look at the inference of partial correlation networks using SPACE, a sparse precision matrix estimator. We find that the networks inferred tend to be quite unstable, and that there is a general tendency for companies

to be connected to those in the same sector. Furthermore we show that the financial, consumer discretionary and industrial sectors tend to be central in these networks.

As previously mentioned, Pearson correlation assumes normality. Financial returns do not tend to follow this assumption, and also contain outliers. Therefore in Chapter 7 we study the use of rank correlation methods, which are robust to these issues. In particular we focus on using Spearman and Kendall's τ in the construction of minimum spanning trees, and how these rank correlation trees compare to those constructed using Pearson correlation. We do find that the rank correlation MSTs tend to maintain more edges over the dataset than the Pearson correlation MSTs, but that most of the instability seems to come from the MST edge selection procedure.

We then investigated how the stability and similarity of correlation networks changes with the volatility of the markets (Chapter 8). To measure this we look at some network measures, notably the mean difference between nodes, the standard deviation of the values for eigenvector centrality and the difference between the eigenvector centralities from adjacent networks. With this we find that the standard deviation and mean difference are negatively correlated with the volatility present in the market, indicating the networks could be considered more similar and stable during times of stress. To further study this we perform some community detection, seeing how the state of the market affects our ability to discover the underlying sector structure, the number of clusters present and the consistency of the clustering. We find that we have less success discovering the underlying sector structure during times of higher volatility and the number of clusters drop but the clustering consistency increases, again indicating the networks could be considered more stable.

9.2 Discussion

Drawing the work together, we have some broad conclusions to make. The first is that the sparse precision matrix estimation methods we have evaluated do not seem to perform well, being outperformed on network inference by a simple thresholding of the sample covariance matrix, by the Ledoit-Wolf methods in portfolio optimization and generally producing very unstable networks when applied to financial data. This is on top of the difficulties of selecting a regularization parameter. Previous work has reflected similar challenges in the use of these methods and we would caution against their application [85] [217].

A second conclusion is that the application of the topological filtration methods can cause the resulting networks to have different properties to that of the full networks. For instance, in chapter 7 we showed that the MST construction method has more of an effect than the choice of correlation coefficient, and in chapter 8 we showed that the PMFG networks do not show the increase in node similarity demonstrated in the full correlation networks. While this may seem obvious, we feel that many authors do not consider the effects that these methods have when they are applied.

This leads us to an issue touched upon in the conclusion of chapter 2, and a big issue in the field of the inference of networks from asset returns, the lack of ground truth models. We do not have much of a reason to believe that the results of any of the filtration methods truly reflect the underlying structure of the markets, and few results to demonstrate their improved performance on real life problems objectively.

This also reflects one of our challenges when attempting to evaluate the different methods that quantify asset relationships. Generally the field tends to evaluate the success of methods by seeing if pre-known events or structure can be seen in the

resulting networks. This is not an objective measure, and therefore it can be difficult to directly compare methods and provide an objective result at the end.

The lack of objective results limits the practical application of these networks in finance. We do feel they show promise for making broad recommendations, for instance showing that correlation increases during times of stress, making diversification challenging [211] [223] (also see our work in Chapter 8). However, these are not specific recommendations (i.e. purchase this particular asset), and we feel these issues mean the networks should not be used for this purpose.

This also ties into how these networks may have an impact for financial practitioners in the medium term. We suggest that using simple methods (i.e. Pearson correlation with no filtration) to assist humans in making portfolio decisions would be the best approach here, as this allows the decision makers to take the downsides into account. Currently we do not believe the filtration methods are likely to provide much of an impact in the short or medium term, due to the many issues discussed. With the poor performance of the sparse precision matrix estimation methods, and the lack of much extractable structure from the partial correlation networks, we also do not feel they will provide much practical use.

9.3 Future Work

As previously mentioned, one of the biggest issues in the field of financial networks is the lack of objective results on the performance of the filtration models. Our first suggestion is therefore to conduct a detailed comparison of the various topological methods, comparing how successfully they retain information while avoiding overfitting, in a similar manner to our chapter comparing correlation coefficients. A possible approach could be to generate datasets which would have different correlation matrices (e.g. those from control patients vs those with cancer), applying the filtration methods and then using a graph classification method to distinguish between the two.

This could be further extended to apply topological filtration methods to data other than financial returns, in the hope to obtain more information on how the methods perform. There has been initial work in the neuroscience [253] [254] and bioinformatics domains [242] [240], but this is only a small amount of work. We would be particularly interested in looking at gene expression data, potentially from The Human Cancer Genome project, where we hope the filtration methods would allow us to understand how genes are differently expressed in tumour cells as opposed to normal ones. We hope that our software package, TopCorr may help in these plans.

On the theme of a lack of objective models, there is the issue of window size selection for the inference of financial networks. It may even be desirable to have a dynamic window as it is doubtful the financial markets only change at certain points. There has been some work in change point detection for correlation [14] and covariance matrices [4], which could be used to dynamically select an appropriate window. We suggest that these methods could be used to improve the analysis of financial correlation networks, by reducing the 'blurring' that occurs between periods with a different underlying correlation matrix. These have only had limited applications in the inference of financial networks [9] so far, and we feel this could be a fruitful direction to take. In particular, chapters 7 and 8 could benefit from such an approach.

Next we discuss some possibilities for precision matrix estimation. In general, we have shown that the sparse precision matrix estimators do not outperform application

specific baselines. Therefore clearly more work is required in these methods to obtain an estimator that can reliably uncover the underlying precision matrix. CLIME may be a solution to this, but we would require faster optimization programs for this to be computationally feasible for many real world problems. Creating software that makes better use of parallelisation or is written in a lower level language than Python may allow the use of CLIME for more realistically sized problems.

Furthermore, for the lasso based methods (or even for thresholding) selecting the regularization parameter is still a challenging problem. We would be interested in using the objective functions from the problems we are attempting to solve to select the regularization parameter, for instance selecting a λ for the graphical lasso that gives us the lowest risk on a portfolio.

Turning to the partial correlation networks, we found them to be very unstable, and generally lacking detectable structure. A potential solution to this would involve looking at dealing with non-linearities in the data, for instance using non-paranormal models [151] or taking a rank correlation approach. A downside here is that this makes the resulting coefficients much less interpretable. Alternatively, the lack of stability could be dealt with by using a temporal constraint on the matrices to reduce edge changes between adjacent time period, e.g. the time varying graphical lasso [97] or TESLA [1].

Appendix A

Analysis of Partial Correlation MSTs and Asset Graphs

A.1 Introduction

In chapter 5 we compared the partial correlation matrices to their correlation counterparts in terms of stability, clustering, how the entries compare and the centrality of various sectors using the full networks with no filtering. In this chapter we change our approach and compare them in terms of consistency, topology and various network measures by using some topological filtration methods. We are particularly interested in using graph based measures after applying some topological filtration in the hope that we can show which parts of the partial correlation matrix are signal, and which parts are noise.

The correlation and partial correlation networks are inferred in a similar manner to before, using the Ledoit-Wolf shrinkage methods to estimate a covariance matrix from a window of stock data, and then using this covariance matrix to obtain the correlation and partial correlation matrices.

A.2 Data

The data we use is downloaded from Yahoo Finance. We use returns from the S&P500 from 2000/03/01 to 2019/10/21. Any company missing more than 10% of its data is removed, and any missing values are filled forwards from the first good value. If the values are missing from the start we backfill from the first good value.

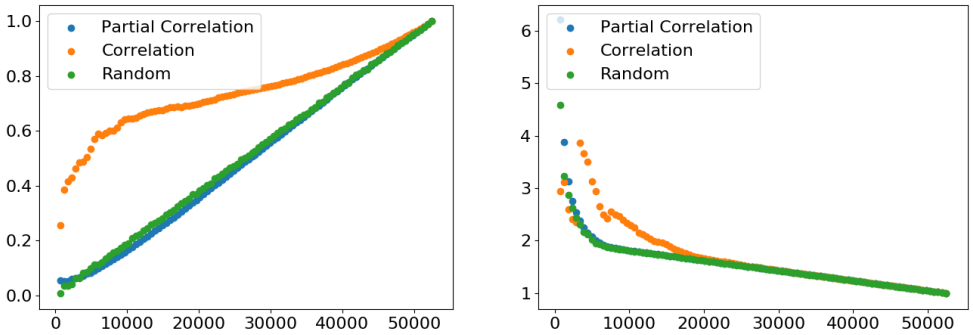
This results in 4790 days of return data for 229 companies. Each company is tagged with a sector from the GICS classification using information from Bloomberg. This places each company into 1 of 11 sectors, Information Technology, Real Estate, Materials, Telecommunication Services, Energy, Financials, Utilities, Industrials, Consumer Discretionary, Healthcare or Consumer Staples.

We make use of Python, NumPy and SciPy [193] for general scripting, pandas [169] for handing the data, statsmodels [227] for some of the statistical analysis, matplotlib [110] for plotting, arch [233] for the implementation of the circular bootstrap and Networkx [96] for the network analysis.

A.3 Results and Analysis

To start with we analyse the first 504 days of data from the S&P500. Taking an approach very similar to that of the 'asset graph' [194] we rank the strength of the off-diagonal edges in the graph from largest to smallest and set a threshold at each of these values. Values above or equal to this threshold are set to 1, those below are

set to 0. With this binary graph we can then study how various network measures change with the number of edges we include. A measure we are particularly interested in is the average clustering coefficient. This is due to evidence that suggests that correlation networks are biased to create 'small world' networks due to the presence of indirect correlations [104] [289]. Of course our partial correlation networks remove this effect, and the authors of these papers do state that in fact partial correlation networks are *less* clustered than would be expected from a random network model, so we are curious to see if these results are reflected in real data. A graph of how the average clustering coefficient changes as we add edges for both a correlation, partial correlation and an Erdos-Renyi random model is shown in Figure A.1 (a).



(a) Average Clustering Coefficient (b) Average Path Length

FIGURE A.1: Clustering Coefficient and Average path length of the correlation, partial correlation and a corresponding random network with the same number of edges.

Effectively the partial correlation network is no more clustered than that of the equivalent random network, with the exception of a small number of values at the start. Onnela et al [194] claim that the point at which the clustering coefficient starts looking like a random graph is the point at which the edges added could be considered noise, which implies that the partial correlation networks might have only a small number of noisy edges.

Having a large average clustering coefficient is not the only measure that makes a network 'small world'. The other measure of interest is the average path length (L). This however is not defined when the network is disconnected, so in these situations we measure the average path length of the largest connected component. A plot of how the average path length changes as we add edges in shown in Figure A.1 (b). Again the partial correlation networks look far more like a random graph than the correlation networks.

We can then quantify the 'small worldness' of a network via the following measure:

$$\sigma = \frac{\frac{C}{C_r}}{\frac{L}{L_r}} \tag{A.1}$$

This is comparing the average clustering coefficient of the graph (C) to that of the equivalent random graph C_r against the average path length in the graph L to the average path length in the equivalent random graph L_r. Again we plot a graph to compare these, this is shown in Figure A.2.

There is a small uptick for the partial correlation values with a very small number of edges, but beyond that it is clearly not a small world network. The correlation

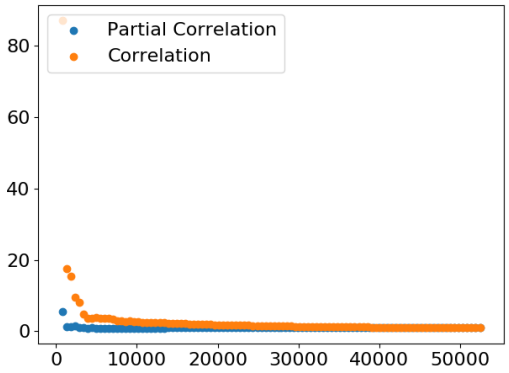


FIGURE A.2: Small World Measure for the two networks

network on the other hand shows a significant amount of small world structure for a large number of edges.

To study this further we turn to an MST based model in a similar manner to Tumminello et al [263]. This allows us to significantly reduce the complexity of the model and to discard potentially noisy edges at the cost of discarding any edges that do not meet our topological constraint. In our experiments we use a circular bootstrap to create a set of pseudo stock returns and calculate the correlation and partial correlation matrices from these. A circular bootstrap is a bootstrap designed for time series data. With a normal bootstrap, we sample randomly with replacement from the dataset paying no attention to the location of a sample. Due to look-ahead effects (i.e. we might gain some information about future performance if we include future samples before ones in the past) this is not a good idea with time series data. In a circular bootstrap we select continuous stretches of time for each pseudo-dataset in order to maintain any time series dependence, and if we reach the end we 'wrap around' the dataset and start back at the first sample (hence the name circular bootstrap).

A bootstrap requires more data than above, so we expand the window to the first 1008 days of data (4 trading years), and set a bootstrap size of 252 days. Once we have these matrices, we can transform them into a distance measure in the manner proposed by Mantegna et al [160]

$$d_{ij} = \sqrt{2 - c_{ij}} \tag{A.2}$$

and create MSTs using Kruskal's algorithm. Our goal is to quantify the stability of these trees - if the partial correlation matrix is mostly noise then we would expect the MSTs to be unstable and be far less consistent in their edge selection than the corresponding correlation matrix. We create 1000 times series using this bootstrap.

To start with we look at the frequency of edge selection. A histogram of the frequency of which edges are selected in shown in Figure A.3. From this we can see that the majority of the edges for both networks are selected a small number of times, and a relatively small number are selected a lot, with the correlation networks selecting a few edges in almost every tree.

Next we look at the differences between the edge selection across the MSTs. To measure this we consider an edge to be present (1) or not (0), and measure the fraction of edge changes that occur between different trees. We calculate this for every tree in the dataset. The mean difference per MST for the correlation is 0.434 ± 0.121 and for the Partial Correlation is 0.555 ± 0.129 indicating the partial

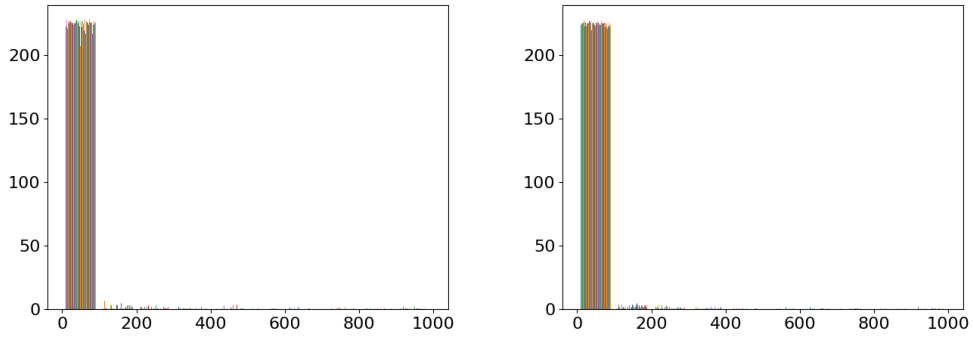


FIGURE A.3: Number of selections of a particular edge in the bootstrapped MSTs. Most edges are selected a small number of times in both MSTs, while a small number are selected many times

correlation MSTs are slightly more unstable than the corresponding correlation ones.

We can get a p-value for the existence of a particular edge by looking at the fraction of trees it is selected in our of the dataset. With this we study how the p-value relates to the strength of the correlation? Previous work has shown there is little relationship between the two in correlation based networks [263]. A scatter plot of these is shown in Figure A.4. It is clear that the larger weights in the partial correlation networks are selected more often than their correlation counterparts (although it is important to remember there are less larger correlations present in these networks) The Spearman correlation between the correlation strength and the p-value is 0.224 for the correlation based MSTs and 0.316 for the partial correlation based MSTs.

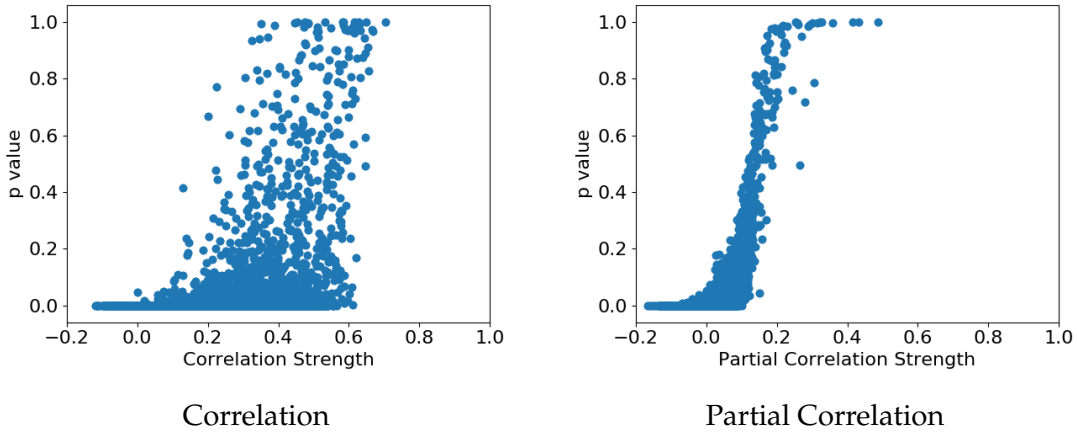
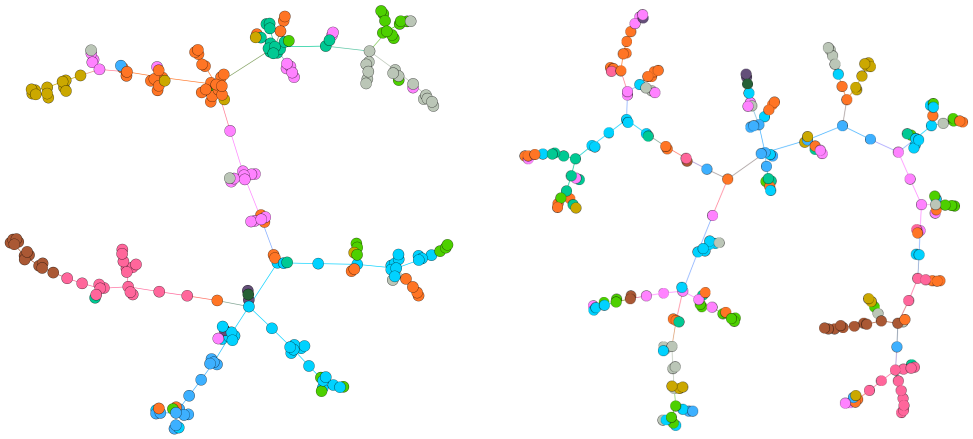


FIGURE A.4: Correlation strength vs p-value for the MSTs.

There is then a question of how the MSTs change over time and how the structure of the partial correlation MSTs varies from the correlation ones. Firstly to explore this we plot example MSTs inferred from the first window of data in Figure A.5. In both we can see a degree of sector clustering, particularly in the correlation MST, with branches tending to be made up of companies in the same sector. This is less the case in the partial correlation MST, with a much larger degree of multisector branches, although companies are still likely to be connected to those in the same sector.



(a) Correlation MST (b) Partial Correlation MST

FIGURE A.5: Example MSTs inferred from the first window using Correlation (a) and Partial Correlation (b)

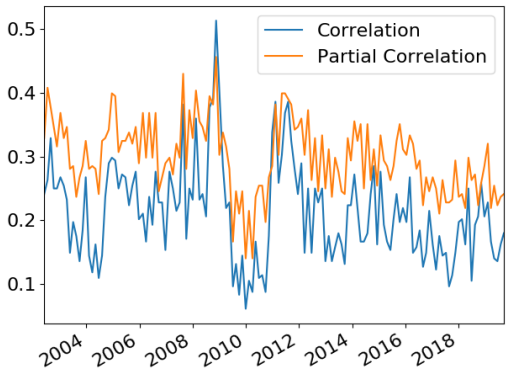


FIGURE A.6: Number of edges changes for the two trees over time. The partial correlation trees have a greater number of edge changes than the correlation MSTs, but both of these are heavily dependent on the market state

To explore how the MSTs change over time we use a sliding window of 504 days, and slide along 30 days at a time for our dataset. We measure the number of edge changes between trees adjacent in time and plot this in Figure A.6. From this we can see that the partial correlation MST maintain significantly less edges between adjacent trees that the correlation MST. What is however interesting is the number of changes for both MSTs is also heavily dependent on the market state, and clearly drops very rapidly for both networks from 2010 - 2012.

For our second measure of stability we look at whether relationships are maintained over time. To do so we plot histograms showing the number of times a particular edge is selected across the dataset. A plot of this is shown in Figure A.7.

The partial correlation MSTs select a greater variety of edges than the correlation MSTs. This would imply they are generally less consistent and could be considered to contain more noise than the correlation ones. What is particularly interesting is that the correlation MSTs select a few edges for almost every single MST inferred through the network, whereas this is not the case for the partial correlation ones, who select an edge for a maximum of half of the dataset.

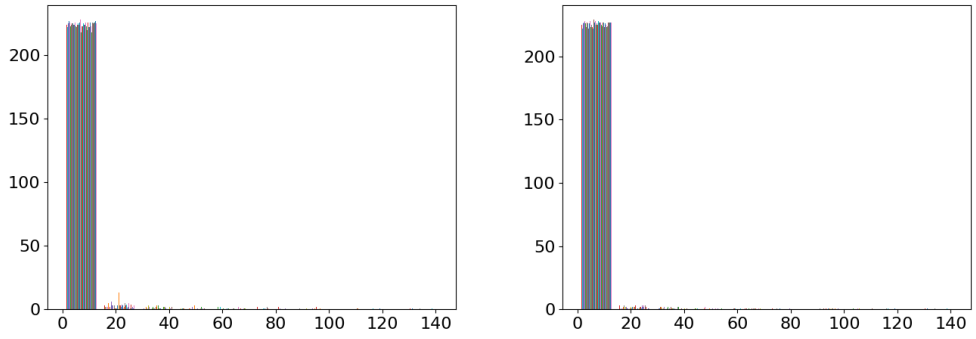


FIGURE A.7: Number of times an edge is selected across the MSTs inferred from the entire dataset. The partial correlation MSTs select a greater variety of edges compared to the correlation MSTs, indicating the network changes more. This could either be an indication there is more noise in the partial correlation networks, or that they are more sensitive to the market state.

A.4 Discussion and Conclusion

In this paper we have compared correlation and partial correlation networks using thresholding and minimum spanning trees. We find there is generally less clustering in the partial correlation networks than the correlation ones, and the structure of the partial correlation networks in general seems to resemble a random network compared to the small world structure of the correlation networks.

Using the minimum spanning tree as a topological filtering mechanism for the networks, we use bootstrapping and study how the trees change over time to see how stable the various networks are. We find that again the trees constructed from the partial correlation networks are significantly less stable than their correlation counterparts. There is however a slightly larger correlation between the relationship strength in the partial correlation network and the number of times an edge is selected in the MST compared to the correlation trees, but for both the correlation is not large.

In conclusion we find these partial correlation networks tend to be noisier than the correlation counterparts. Possible reasons for this could be due to the non-Gaussianity of financial returns, since partial correlation measures linear relationships.

Appendix B

Keyphrase Extraction Using Regularized Covariance Graphs

B.1 Introduction

As the number of documents available increases explosively as time goes on, extracting and summarizing key points becomes ever more critical to assist humans in indexing, organizing and deciding what to read. Keyphrase extraction is a tool that can assist here, where we extract a few words or short phrases that accurately summarize a document. Applications of keyphrase extraction include phishing detection [146], clustering [98] and content recommendation [212]. Extracting keyphrases can be done either in a supervised or unsupervised manner. Supervised methods tend to approach keyphrase extraction as a binary classification problem - is a certain word or phrase a keyphrase? Supervised methods include KEA [83], which uses a Naive Bayes classifier to assign keyphrases, GenEx [265] which uses a genetic algorithm to tune a heuristic method based upon 12 statistical features and [109], which uses a rule based algorithm to output keyphrases. Supervised methods can produce better results than unsupervised methods, but do require large annotated corpora [132]. In many situations training data may be time consuming to construct or simply not available. Therefore we often turn to unsupervised methods to avoid obtaining pre-trained corpora. For this reason, we focus on unsupervised methods in this paper.

Graph based methods can provide us with techniques for unsupervised keyphrase extraction. We construct a graph from words or phrases in the text, using said words or phrases units as nodes and relationships between them as edges. By using various measures of node importance (network centrality) we can decide which units are keyphrases. Previous methods such as RAKE [218] and TextRank [173] adopt this approach, constructing graphs with words as nodes, and creating an edge between two nodes if they exist within a certain distance of each other. TextRank uses PageRank [43] to compute node centrality. PositionRank [82] is another variation on this theme. In this method they construct a word co-occurrence graph and use a biased PageRank method to decide on node centralities. Each word is biased using the reciprocal of it's position in a document - words that occur earlier are given more weight. In a similar manner, Bougounin et al [34] propose TopicRank. Here the authors use a hierarchical clustering algorithm to cluster the keyphrases in the document into topics. They select the most representative phrase from each topic and construct a complete graph using each topic as a node, and the weight on an edge as the reciprocal of the distance between when 2 topics occur in a document.

Statistical methods are an alternative unsupervised approach. The simplest method is TF-IDF, where we calculate how much information a word gives us in a document. Words that are rare in the entire corpus but common in a document are given higher

scores. However TF-IDF does require a corpus. Matsuo and Ishizuka [165] use the χ^2 measure to detect keyphrases. They evaluate their method by showing paper authors the top 15 keywords extracted by their method and TF-IDF and ask them to decide which words are relevant. Performance was similar to TF-IDF, with a precision of 0.51 for their method vs 0.53 for TF-IDF but without the corpus requirement. A more recently proposed method is YAKE!, developed by Campos et al [48]. YAKE! takes 5 features for each word and heuristically combines them to give a score. The candidate keyphrases are then ranked, with a smaller score indicating a more meaningful keyphrase.

We propose to construct covariance graphs from text data and use nodes of a high centrality as keyphrases. Since the covariance matrix is a symmetric matrix, we can treat it like an undirected graph, with each dimension as a node and the covariance between two dimensions being the weight on the edge. Such an approach has been used before in financial applications [26] [175], biology [141] and neuroscience [244]. To our knowledge this has not been directly investigated before with text data. LexRank [75] uses cosine similarity (which is very similar Pearson correlation, which can be thought of as normalized covariance) to evaluate how important a sentence is in automated summarization, but they do not attempt to extract keyphrases. Li and Han [148] use correlation and PCA to extract keyphrase. They weight each potential keyphrase using TF-IDF, word length, word co-occurrence and the part of speech and calculate the correlation between each word based upon these factors. From this correlation matrix they extract the eigenvalues and use each eigenvalue to weight it's respective factor. These weighted factors are then summed up to produce a score for the keyphrase. Bohne and Borghoff [28] use PCA to decide which factors are important in keyphrase extraction. They find that combining statistical techniques, the Γ-Metric (developed by [292]) TF-IDF, and the Bernoulli model gives better results than the methods alone with regards to the precision of the keyphrases extracted.

B.2 Covariance Graph Construction

To begin with we must turn our document into a matrix in order to calculate a covariance matrix. To start with we search through the document for candidate keyphrases. This is common in other methods, as generally keyphrases are likely to include nouns. We use the same candidate method as in PositionRank [82], considering phrases whose part of speech tags match the regular expression (adjective)*(noun)+, using phrases up to a length of 3. We split the document into sentences, count the number each ngrams in each sentence and use this for a row of the matrix. Constructing this for each sentence into the document we produce a n by p matrix X, where n is the number of sentences and p is the number of words.

The maximum likelihood estimator for a covariance matrix (also known as the sample covariance matrix) can be estimated using

$$S = \frac{1}{n}\sum_{i=1}^{N}(\boldsymbol{x_i} - \boldsymbol{\mu})^T(\boldsymbol{x_i} - \boldsymbol{\mu}) \tag{B.1}$$

where $\boldsymbol{x_i}$ is a row of the word occurrence matrix X and $\boldsymbol{\mu}$ is the mean of the words occurrence. In our case, we do not center the data when calculating the covariance

matrix.

$$S = \frac{1}{n}\sum_{i=1}^{N} \boldsymbol{x}_i^T(\boldsymbol{x}_i) \tag{B.2}$$

This makes our covariance matrix non-negative. Non-negative matrices have a largest eigenvalue which is positive and unique, with the corresponding leading eigenvector having positive entries. This is important when interpreting each entry as a node centrality measure.

Text data is famous for being high-dimensional (i.e. more dimensions than samples) and it is well known that in these situations, the covariance matrix can be ill-formed [142]. Therefore is it common in many fields to apply regularization to the covariance matrix [86] [225] [53]. In this particular case we use shrinkage, combining the sample covariance matrix with a well-regularized matrix, the identity matrix (I)

$$S^* = \rho S + (1-\rho)I \tag{B.3}$$

The regularization parameter ρ can be calculated analytically. Ledoit and Wolf [142] propose to select it by minimizing the squared error between the true covariance matrix and the sample covariance matrix. This is extended by Chen et al [53], who show that if Gaussianity of the data is assuming we can calculate a more optimal value for ρ. We utilize both of these in this paper to see if shrinkage can improve performance.

We then must decide which nodes are the most central. To do so we experiment with 2 centrality measures. The first measure is eigenvector centrality, where the leading eigenvector is extracted from the covariance matrix (referred to in the results as Unbiased Covariance). Extracting a leading eigenvector is very similar to PCA, a common technique used to reduce the dimensionality of data in numerous fields. Latent Semantic Analysis (LSA) [140] is possibly the most famous application of PCA in text. Extracting this leading eigenvector is therefore the same as taking 1 dimension with PCA.

Secondly we look at the possibility of including the position of words into our centrality measures, as it has been observed in many previous papers that words that occur closer to the start of a document are more likely to be keyphrases [62] [82] [48]. To do so we use a biased PageRank method in a similar manner to PositionRank [82]. First we describe the PageRank [43] algorithm. Let G be our undirected graph with the ngrams as it's nodes and the correlation between the ngrams as the weights. A is the adjacency matrix corresponding to the graph. We compute the PageRank by a power method, extracting the leading eigenvector $\boldsymbol{s}$ from A using the following equation

$$\boldsymbol{s}(t+1) = \hat{A}.\boldsymbol{s}(t) \tag{B.4}$$

where $\hat{A}$ is the column normalized matrix of A (i.e. the sum of each column of $\hat{A} = 1$. We initialize $\boldsymbol{s}$ to $\frac{1}{p}$ where p is the number of ngrams. The idea behind PageRank is that it is a user randomly wandering around the graph, travelling along edges from one node to the next. To ensure the user does not get stuck, we can add a 'teleporting' factor - i.e. the user jumps to another node with change α. Our computation then becomes

$$\boldsymbol{s}(t+1) = \alpha\hat{A}.\boldsymbol{s}(t) + \frac{(1-\alpha)}{p}\mathbf{1} \tag{B.5}$$

However, we can bias the PageRank algorithm and make the user more likely to walk to certain nodes by using a different vector to $\mathbf{1}$. In PositionRank [82] the authors bias the user towards candidates that occur closer to the start of the

document using a biased PageRank method. We also do this. We experimented with 2 different ways of weighting each ngram, the first with it's inverse position in the document, summing all the position weights if the word occurs multiple times (referred to as Inverse Bias). With this weighting, we found that the method tended towards simply outputting a combination of ngrams found in the first sentence of the document. Therefore we experimented with a second weighting, where the first position in the document is equal to 1 and the second position is equal to $\frac{(\text{num words}-1)}{\text{num words}}$ and so on (referred to as Linear Bias).

We sum up the steps involved below

1. Scan document for candidate phrases
2. Split document into sentences
3. Count occurrence of each potential candidate in each sentence to form word occurrence matrix X
4. Calculate covariance matrix $C = X^T X$
5. Apply shrinkage if desired
6. Calculate weighting of words by summing up position weights
7. Use a centrality measure (either eigenvector centrality or a biased Page Rank algorithm) to rank nodes on importance

B.3 Methods & Data

For each document we remove stop words using the NLTK stopword list and stem each word using the Snowball Stemmer. The datasets we use to evaluate our method are compiled by Wan et al [272] (DUC2004 annotated news articles) and a dataset provided by Nguyen and Kan [186] consisting of research papers from various disciplines. We remove the abstract and copyright information, and use the remaining text as input into the keyphrase extraction methods. This dataset is quite messy due to the papers being turned into plain text files - there are many figure and mathematical artficats in the documents that make them challenging to parse. All these datasets come pre-annotated with keyphrases, and any documents that do not have pre-annotated keyphrases are discarded. Both the true keyphrases and the extracted keyphrases are stemmed with the Snowball Stemmer.

The methods are written in Python 3 using numpy and scipy. We rely on sklearn to provide the implementation of the vectorizing methods, pke [33] for the implementation of TopicRank, PositionRank, TextRank and TFIDF, and rake-nltk for the implementation of RAKE. We will provide the code for this paper on Github.

B.4 Results

Evaluating keyphrase extraction methods is challenging due to the lack of an underlying ground truth - deciding which phrases represent a document is fundamentally a subjective task. We start by evaluating the various methods by comparing the F1 score of the top 5, 10 and 15 candidates. The results are shown in Table B.1.

Method	5		10		15	
	F_1	Precision	F_1	Precision	F_1	Precision
Covariance Linear Bias	21.042 ± 14.98	27.172 ± 18.89	23.857 ± 12.84	21.526 ± 11.79	23.752 ± 11.27	18.321 ± 9.01
Unbiased Covariance	9.266 ± 11.31	11.650 ± 14.03	12.155 ± 11.11	10.748 ± 9.72	13.518 ± 10.37	10.261 ± 7.85
LW Covariance	19.770 ± 15.81	25.421 ± 19.81	23.999 ± 13.65	21.627 ± 12.48	24.001 ± 11.57	18.501 ± 9.20
OAS Covariance	19.914 ± 15.68	25.690 ± 19.74	24.132 ± 13.37	21.762 ± 12.25	24.041 ± 11.55	18.546 ± 9.24
TextRank	9.120 ± 11.36	11.919 ± 14.93	13.693 ± 12.70	12.357 ± 11.66	15.038 ± 11.64	11.627 ± 9.21
RAKE	1.644 ± 4.79	2.155 ± 6.20	3.391 ± 6.29	3.064 ± 5.66	4.233 ± 6.55	3.258 ± 4.99
TopicRank	18.413 ± 15.80	23.569 ± 19.95	19.745 ± 13.34	17.710 ± 11.95	18.732 ± 11.42	14.388 ± 8.87
PositionRank	21.828 ± 15.03	28.199 ± 19.16	24.897 ± 13.83	22.469 ± 12.68	24.449 ± 11.81	18.838 ± 9.34
TFIDF	9.373 ± 11.67	11.785 ± 14.49	10.697 ± 10.18	9.495 ± 8.96	10.554 ± 8.80	8.013 ± 6.61
YAKE	9.327 ± 11.60	12.088 ± 15.08	10.863 ± 10.12	9.787 ± 9.12	11.480 ± 8.98	8.822 ± 6.90

Method	5		10		15	
	F_1	Precision	F_1	Precision	F_1	Precision
Covariance Linear Bias	2.817 ± 5.62	4.929 ± 9.66	5.502 ± 7.05	6.114 ± 8.15	6.507 ± 6.66	6.008 ± 6.54
Unbiased Covariance	6.967 ± 10.30	11.469 ± 17.29	10.242 ± 10.77	11.043 ± 11.96	11.657 ± 9.98	10.458 ± 9.36
LW Covariance	1.407 ± 4.23	2.559 ± 7.73	2.678 ± 5.63	3.081 ± 6.64	3.847 ± 5.84	3.575 ± 5.79
OAS Covariance	1.470 ± 4.31	2.654 ± 7.82	2.854 ± 5.75	3.270 ± 6.76	4.027 ± 6.14	3.733 ± 6.03
TextRank	1.263 ± 3.74	2.275 ± 6.35	2.440 ± 4.34	2.844 ± 5.29	2.947 ± 4.59	2.780 ± 4.48
RAKE	0.126 ± 1.45	0.190 ± 1.94	0.179 ± 1.36	0.200 ± 1.44	0.144 ± 1.07	0.133 ± 0.96
TopicRank	9.903 ± 11.65	15.829 ± 18.41	11.987 ± 10.73	12.991 ± 11.89	11.568 ± 9.49	10.497 ± 9.07
PositionRank	5.257 ± 7.70	9.005 ± 13.08	6.712 ± 7.25	7.583 ± 8.34	7.885 ± 7.12	7.337 ± 7.22
TFIDF	13.629 ± 13.07	21.801 ± 20.20	15.955 ± 11.85	17.251 ± 13.10	16.010 ± 10.62	14.412 ± 10.26
YAKE	11.920 ± 11.52	19.550 ± 18.30	14.598 ± 10.52	16.106 ± 12.06	15.114 ± 9.69	13.838 ± 9.42

TABLE B.1: F1 and precision % score for the results of keyphrase extraction on the Wan dataset (top) and Nguyen dataset (bottom). We separate out our covariance based methods (top section), word occurrence graph based methods (middle section) and statistical methods (bottom section).

To check the sensitivity of the method to the value of α we experimented with setting α to values in the range 0.7, 0.8 and 0.9 to see how this would affect results. We did not find this varied the results much, with around a 1% point difference in the F_1 scores and so we set α to 0.8.

We show the mean and standard deviation of the F_1 score and precision for the Wan dataset (top) and Nguyen dataset (bottom) in Table B.1. It is worth noting the large standard deviation values that make objective assessment of the success of these methods challenging. We only experiment with the shrunk covariance methods using the linear position bias.

Firstly we focus on the Wan dataset. Here we can see our method does not outperform the state of the art (PositionRank) when measured by the F1 score, but that including position very much increases performance in both the covariance (Covariance Linear Bias) and word-occurrence (PositionRank) methods. This could be due to the Wan dataset being news articles - often many of the keyphrases are found in the first sentence. On this dataset graph based methods show better performance than the statistical methods.

The Nguyen dataset shows a very different set of results. Here we can see that graph based methods are outperformed by the statistical methods, and that of particular note, including position information lowers performance of the graph based methods. This could be due to the scientific articles being longer and containing more topics than the news based articles - all of the information contained in the document is not neatly summarized in the first few sentences, possibly explaining why TopicRank has the highest performance. Our unbiased covariance method also has reasonable performance on this dataset, particularly as we start to extract a larger number of keyphrases it's performance is similar to that of TopicRank.

Our second result is that shrinkage provides a small boost to the performance of the method on the Wan dataset, but it is not drastic, and reduces performance on the Nguyen dataset. This is perhaps surprising given the high-dimensional nature of the dataset, but it is also important to pay attention to the conditions required by the shrinkage methods. OAS assumes Gaussianity of the data, and text data is clearly not Gaussian. Therefore it is perhaps less surprising to see it contributes relatively little to improving the performance. The Ledoit-Wolf estimator is much less restrictive in it's assumptions but does require that the distribution of the data has finite 4th moments (kurtosis). Word occurrence data is heavy tailed, and it is possible that it does not have finite kurtosis.

Finally we provide an example of the keywords extracted from the article. This is shown in Table B.2. The news article is on drought in North Dakota. We can see that the covariance method with a linear bias captures much of the various topics in the document.

B.5 Conclusion

In this paper we proposed a method for keyphrase extraction based upon constructing a covariance matrix from a word occurrence matrix and ranking the importance of each word or phrase using various centrality measures. We also investigated the use of shrinkage in this situation, experimenting with both Ledoit-Wolf and OAS shrinkage. We find that in certain situations the performance of our method is comparable to word occurrence graph based keyphrase extraction methods. Manually examining the keyphrases indicates that in many cases important words are extracted, with a fair spread across the topics in the document. We also find that shrinkage is

Method	Keywords
True Keywords	drought,upper midwest,north dakota,disaster relief aid,four-part series,disaster aid program,farmers
Unbiased Covariance	North Dakota,North,Dakota,drought aid,most disaster dollars,drought,drought relief program,Dust Bowl,barren horizon,tumbleweed dancing,aid,economic situation,business closures,areas Congress,their land
Covariance Linear	disaster relief aid,drought aid,disaster aid program,drought relief program,other state,Drought payments,disaster aid,North Dakota,average farmer,many farmers,Midwest drought area,most farmers,drought,disaster relief program,most disaster dollars
LW Covariance	disaster relief aid,Drought payments,other state,drought aid,disaster aid program,drought relief program,average farmer,disaster aid,many farmers,North Dakota,Financial aid,Midwest drought area,most farmers,disaster relief program,drought
OAS Covariance	disaster relief aid,drought aid,Drought payments,disaster aid program,other state,drought relief program,disaster aid,North Dakota,average farmer,many farmers,Midwest drought area,Financial aid,most farmers,drought,disaster relief program
TextRank	disaster relief aid,dakota state university study,iowa state university study,red last year,north dakota state economist,disaster aid program,most disaster dollars,government last year,drought aid,drought relief program,midwest drought area,red river valley county,last thing north dakota,disaster aid,disaster relief program
RAKE	start coming around,shriveled brown stumps,reach areas congress,perhaps nowhere harder,fading small towns,dust bowl days,banker says go,major new wave,farm dollar triples,assistance allocated nationwide,12 months ago,election year ...,drought evoked memories,000 producers shared,000 maximum payment
TopicRank	farmers,drought,north dakota,disaster relief aid,federal aid,years,state,carl grindberg,red spring wheat,crop insurance,crop loss,golden acres,crop,normal trials,others
PositionRank	drought aid,midwest drought area,drought relief program,disaster relief aid,disaster aid program,drought payments,drought assistance,disaster aid,north dakota,drought,disaster relief program,most disaster dollars,other state,financial aid,government aid
YAKE	north dakota,north,dakota,said,perhaps nowhere harder,aid,north dakota state,drought,disaster,state,crop,year,farmers,grindberg,dakota state
TF-IDF	drought,crop,farmers,dakota,north dakota,north,disaster,aid,farm,grindberg,acres,1988,the drought,disaster aid,wheat

TABLE B.2: Top 15 candidate keyphrases extracted by each method.

not particularly useful in this context, while it provides a small boost in performance on one dataset, it significantly reduces it on the other. Future work could involve trying covariance matrices regularized in a different way or using a different method to decide upon when to take sample boundaries - in this paper we use sentence boundaries, but many of the word occurrence graph based methods have had success with setting a co-occurrence window.

Bibliography

[1] Ahmed, Amr and Xing, Eric P. "Recovering time-varying networks of dependencies in social and biological studies". In: *Proceedings of the National Academy of Sciences* 106.29 (2009), pp. 11878–11883.

[2] Akaike, H. "A new look at the statistical model identification". In: *IEEE Transactions on Automatic Control* 19.6 (1974), pp. 716–723. ISSN: 2334-3303. DOI: 10.1109/TAC.1974.1100705.

[3] Albert, Réka, Jeong, Hawoong, and Barabási, Albert-László. "Error and attack tolerance of complex networks". In: *nature* 406.6794 (2000), p. 378.

[4] Aminikhanghahi, Samaneh and Cook, Diane J. "A survey of methods for time series change point detection". In: *Knowledge and information systems* 51.2 (2017), pp. 339–367. DOI: 10.1007/s10115-016-0987-z.

[5] Aste, Tomaso, Shaw, William, and Di Matteo, Tiziana. "Correlation structure and dynamics in volatile markets". In: *New Journal of Physics* 12.8 (2010), p. 085009.

[6] Awoye, Oluwatoyin Abimbola. "Markowitz Minimum Variance Portfolio Optimization using New Machine Learning Methods". PhD thesis. (UCL) University College London, 2016.

[7] Bahaludin, Hafizah, Abdullah, Mimi Hafizah, and Salleh, Supian Mat. "Minimal spanning tree for 100 companies in Bursa Malaysia". In: *AIP Conference Proceedings* 1643.1 (2015), pp. 609–615. DOI: 10.1063/1.4907501. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4907501. URL: https://aip.scitation.org/doi/abs/10.1063/1.4907501.

[8] Balcı, Mehmet Ali, Akgüller, Ömer, and Güzel, Serdar Can. "Hierarchies in communities of UK stock market from the perspective of Brexit". In: *Journal of Applied Statistics* 0.0 (2020), pp. 1–19. DOI: 10.1080/02664763.2020.1796942. eprint: https://doi.org/10.1080/02664763.2020.1796942. URL: https://doi.org/10.1080/02664763.2020.1796942.

[9] Banerjee, Sayantan and Guhathakurta, Kousik. "Change-point analysis in financial networks". In: *Stat* 9.1 (2020). e269 sta4.269, e269. DOI: https://doi.org/10.1002/sta4.269. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sta4.269. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sta4.269.

[10] Barabási, Albert-László. *Linked: The new science of networks*. 2003.

[11] Barabási, Albert-László. "Network science". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1987 (2013), p. 20120375.

[12] Barabási, Albert-László and Albert, Réka. "Emergence of Scaling in Random Networks". In: *Science* 286.5439 (1999), pp. 509–512.

[13] Barfuss, Wolfram et al. "Parsimonious modeling with information filtering networks". In: *Phys. Rev. E* 94 (6 2016), p. 062306. DOI: 10.1103/PhysRevE.94.062306. URL: https://link.aps.org/doi/10.1103/PhysRevE.94.062306.

[14] Barnett, Ian and Onnela, Jukka-Pekka. "Change point detection in correlation networks". In: *Scientific reports* 6 (2016), p. 18893.

[15] Baruchi, Itay, Towle, Vernon L., and Ben-Jacob, Eshel. "Functional holography of complex networks activity—From cultures to the human brain". In: *Complexity* 10.3 (2005), pp. 38–51. DOI: 10.1002/cplx.20065. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cplx.20065. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cplx.20065.

[16] Bastian, Mathieu, Heymann, Sebastien, and Jacomy, Mathieu. "Gephi: an open source software for exploring and manipulating networks". In: *Third international AAAI conference on weblogs and social media*. 2009.

[17] Begušić, Stjepan et al. "Information Feedback in Temporal Networks as a Predictor of Market Crashes". In: *Complexity* 2018 (2018).

[18] Bell, F.K. "A note on the irregularity of graphs". In: *Linear Algebra and its Applications* 161 (1992), pp. 45 –54. ISSN: 0024-3795. DOI: https://doi.org/10.1016/0024-3795(92)90004-T. URL: http://www.sciencedirect.com/science/article/pii/002437959290004T.

[19] BELLONI, A., CHERNOZHUKOV, V., and WANG, L. "Square-root lasso: pivotal recovery of sparse signals via conic programming". In: *Biometrika* 98.4 (2011), pp. 791–806. ISSN: 00063444. URL: http://www.jstor.org/stable/23076172.

[20] Ben-Hur, Asa and Guyon, Isabelle. "Detecting stable clusters using principal component analysis". In: *Functional genomics*. Springer, 2003, pp. 159–182.

[21] Best, Michael J and Grauer, Robert R. "On the sensitivity of mean-variance-efficient portfolios to changes in asset means: some analytical and computational results". In: *Review of Financial Studies* 4.2 (1991), pp. 315–342.

[22] Bhattacharjee, Biplab, Shafi, Muhammad, and Acharjee, Animesh. "Network mining based elucidation of the dynamics of cross-market clustering and connectedness in Asian region: An MST and hierarchical clustering approach". In: *Journal of King Saud University - Computer and Information Sciences* 31.2 (2019), pp. 218 –228. ISSN: 1319-1578. DOI: https://doi.org/10.1016/j.jksuci.2017.11.002. URL: http://www.sciencedirect.com/science/article/pii/S1319157817302343.

[23] Billio, Monica et al. "Econometric measures of connectedness and systemic risk in the finance and insurance sectors". In: *Journal of Financial Economics* 104.3 (2012). Market Institutions, Financial Market Risks and Financial Crisis, pp. 535 –559. ISSN: 0304-405X. DOI: https://doi.org/10.1016/j.jfineco.2011.12.010. URL: http://www.sciencedirect.com/science/article/pii/S0304405X11002868.

[24] Birch, Jenna, Pantelous, Athanasios A., and Soramäki, Kimmo. "Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns". In: *Computational Economics* 47.4 (2016), pp. 501–525. ISSN: 1572-9974. DOI: 10.1007/s10614-015-9481-z. URL: https://doi.org/10.1007/s10614-015-9481-z.

[25] Blondel, Vincent D et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[26] Boginski, Vladimir, Butenko, Sergiy, and Pardalos, Panos M. "Statistical analysis of financial networks". In: *Computational Statistics & Data Analysis* 48.2 (2005), pp. 431 –443. DOI: 10.1016/j.csda.2004.02.004.

[27] Boginski, Vladimir et al. "A network-based data mining approach to portfolio selection via weighted clique relaxations". In: *Annals of Operations Research* 216.1 (2014), pp. 23–34.

[28] Bohne, T. and Borghoff, U. M. "Data Fusion: Boosting Performance in Keyword Extraction". In: *2013 20th IEEE International Conference and Workshops on Engineering of Computer Based Systems (ECBS)*. 2013, pp. 166–173. DOI: 10.1109/ECBS.2013.12.

[29] Bonanno, Giovanni, Vandewalle, Nicolas, and Mantegna, Rosario N. "Taxonomy of stock market indices". In: *Physical Review E* 62.6 (2000), R7615.

[30] Bonanno, Giovanni et al. "Topology of correlation-based minimal spanning trees in real and model markets". In: *Phys. Rev. E* 68 (4 2003), p. 046130. DOI: 10.1103/PhysRevE.68.046130. URL: https://link.aps.org/doi/10.1103/PhysRevE.68.046130.

[31] Borghesi, Christian, Marsili, Matteo, and Micciche, Salvatore. "Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode". In: *Physical Review E* 76.2 (2007), p. 026104.

[32] Bouchaud, Jean-Philippe and Potters, Marc. "Financial applications of random matrix theory: a short review". In: *arXiv preprint arXiv:0910.1205* (2009).

[33] Boudin, Florian. "pke: an open source python-based keyphrase extraction toolkit". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*. Osaka, Japan, 2016, pp. 69–73. URL: http://aclweb.org/anthology/C16-2015.

[34] Bougouin, Adrien, Boudin, Florian, and Daille, Béatrice. "Topicrank: Graph-based topic ranking for keyphrase extraction". In: *International Joint Conference on Natural Language Processing (IJCNLP)*. 2011, pp. 543–551.

[35] Bradley Efron, R.J. Tibshirani. *An introduction to bootstrap*. 1st ed. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman & Hall, 1994. ISBN: 9780412042317,0-412-04231-2. URL: http://gen.lib.rus.ec/book/index.php?md5=BF16596F119E917AD6501BF994E30C9E.

[36] Brandes, Ulrik et al. "Maximizing modularity is hard". In: *arXiv preprint physics/0608255* (2006).

[37] Brida, J. Gabriel and Risso, W. Adrián. "Hierarchical structure of the German stock market". In: *Expert Systems with Applications* 37.5 (2010), pp. 3846 –3852. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2009.11.034. URL: http://www.sciencedirect.com/science/article/pii/S0957417409009762.

[38] Brida, Juan Gabriel, Gómez, David Matesanz, and Risso, Wiston Adrián. "Dynamical hierarchical tree in currency markets". In: *Documentos de Trabajo FUNCAS* 332 (2007), p. 1.

[39] Brida, Juan Gabriel, Matesanz, David, and Seijas, Maria Nela. "Network analysis of returns and volume trading in stock markets: The Euro Stoxx case". In: *Physica A: Statistical Mechanics and its Applications* 444 (2016), pp. 751 –764. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.10.078. URL: http://www.sciencedirect.com/science/article/pii/S0378437115009371.

[40] Brida, Juan Gabriel and Risso, W. Adrian. "Dynamic and Structure of the Italian stock market based on returns and volume trading". In: *Economics Bulletin* 29.3 (2009), pp. 2417–2423. URL: https://ideas.repec.org/a/ebl/ecbull/eb-09-00306.html.

[41] BRIDA, JUAN GABRIEL and RISSO, WISTON ADRIAN. "DYNAMICS AND STRUCTURE OF THE MAIN ITALIAN COMPANIES". In: *International Journal of Modern Physics C* 18.11 (2007), pp. 1783–1793. DOI: 10.1142/S0129183107011741. eprint: https://doi.org/10.1142/S0129183107011741. URL: https://doi.org/10.1142/S0129183107011741.

[42] Brier, Matthew R. et al. "Partial covariance based functional connectivity computation using Ledoit–Wolf covariance regularization". In: *NeuroImage* 121 (2015), pp. 29–38. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2015.07.039. URL: http://www.sciencedirect.com/science/article/pii/S1053811915006503.

[43] Brin, Sergey and Page, Lawrence. "The Anatomy of a Large-Scale Hypertextual Web Search Engine". In: *COMPUTER NETWORKS AND ISDN SYSTEMS*. 1998, pp. 107–117.

[44] Brodie, Joshua et al. "Sparse and stable Markowitz portfolios." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.30 (2009), pp. 12267–72.

[45] Bun, Joël, Bouchaud, Jean-Philippe, and Potters, Marc. *My beautiful laundrette: Cleaning correlation matrices for portfolio optimization*. 2016.

[46] Burdekin, Richard C. K., Hughson, Eric, and Gu, Jinlin. "A first look at Brexit and global equity markets". In: *Applied Economics Letters* 25.2 (2018), pp. 136–140. DOI: 10.1080/13504851.2017.1302057. eprint: https://doi.org/10.1080/13504851.2017.1302057. URL: https://doi.org/10.1080/13504851.2017.1302057.

[47] Cai, Tony, Liu, Weidong, and Luo, Xi. "A Constrained l1 Minimization Approach to Sparse Precision Matrix Estimation". In: *Journal of the American Statistical Association* 106.494 (2011), pp. 594–607. DOI: 10.1198/jasa.2011.tm10155. eprint: https://doi.org/10.1198/jasa.2011.tm10155. URL: https://doi.org/10.1198/jasa.2011.tm10155.

[48] Campos, Ricardo et al. "YAKE! collection-independent automatic keyword extractor". In: *European Conference on Information Retrieval*. Springer. 2018, pp. 806–810.

[49] Candes, Emmanuel and Tao, Terence. "The Dantzig selector: Statistical estimation when p is much larger than n". In: *Ann. Statist.* 35.6 (Dec. 2007), pp. 2313–2351. DOI: 10.1214/009053606000001523. URL: https://doi.org/10.1214/009053606000001523.

[50] Candès, Emmanuel J., Romberg, Justin K., and Tao, Terence. "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on Pure and Applied Mathematics* 59.8 (2006), pp. 1207–1223. DOI: 10.1002/cpa.20124. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.20124. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20124.

[51] Cao, Guangxi, Xu, Longbing, and Cao, Jie. "Multifractal detrended cross-correlations between the Chinese exchange market and stock market". In: *Physica A: Statistical Mechanics and its Applications* 391.20 (2012), pp. 4855 –4866. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.05.035. URL: http://www.sciencedirect.com/science/article/pii/S0378437112004062.

[52] Care, Matthew A, Westhead, David R, and Tooze, Reuben M. "Parsimonious Gene Correlation Network Analysis (PGCNA): a tool to define modular gene co-expression for refined molecular stratification in cancer". In: *NPJ systems biology and applications* 5.1 (2019), pp. 1–17.

[53] Chen, Yilun et al. "Shrinkage algorithms for MMSE covariance estimation". In: *IEEE Transactions on Signal Processing* 58.10 (2010), pp. 5016–5029.

[54] Christensen, Alexander P. "NetworkToolbox: Methods and Measures for Brain, Cognitive, and Psychometric Network Analysis in R". In: *The R Journal* 10.2 (2018), pp. 422–439. DOI: 10.32614/RJ-2018-065. URL: https://doi.org/10.32614/RJ-2018-065.

[55] Chu, J. and Nadarajah, S. "A statistical analysis of UK financial networks". In: *Physica A: Statistical Mechanics and its Applications* 471.Supplement C (2017), pp. 445 –459.

[56] Coelho, R. et al. "Sector analysis for a FTSE portfolio of stocks". In: *Physica A: Statistical Mechanics and its Applications* 373 (2007), pp. 615 –626. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2006.02.050. URL: http://www.sciencedirect.com/science/article/pii/S0378437106006364.

[57] Coelho, Ricardo et al. "The evolution of interdependence in world equity markets—Evidence from minimum spanning trees". In: *Physica A: Statistical Mechanics and its Applications* 376 (2007), pp. 455 –466. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2006.10.045. URL: http://www.sciencedirect.com/science/article/pii/S0378437106010624.

[58] Coletti, Paolo. "Comparing minimum spanning trees of the Italian stock market using returns and volumes". In: *Physica A: Statistical Mechanics and its Applications* 463 (2016), pp. 246 –261. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2016.07.029. URL: http://www.sciencedirect.com/science/article/pii/S0378437116304605.

[59] Cont, R. "Empirical properties of asset returns: stylized facts and statistical issues". In: *Quantitative Finance* 1.2 (2001), pp. 223–236. DOI: 10.1080/713665670. eprint: https://doi.org/10.1080/713665670. URL: https://doi.org/10.1080/713665670.

[60] Coronnello, C et al. "Sector identification in a set of stock return time series traded at the London Stock Exchange". In: *arXiv preprint cond-mat/0508122* (2005).

[61] Curme, Chester et al. "Emergence of statistically validated financial intraday lead-lag relationships". In: *Quantitative Finance* 15.8 (2015), pp. 1375–1386. DOI: 10.1080/14697688.2015.1032545. eprint: https://doi.org/10.1080/14697688.2015.1032545. URL: https://doi.org/10.1080/14697688.2015.1032545.

[62] Danesh, Soheil, Sumner, Tamara, and Martin, James H. "Sgrank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction". In: *Proceedings of the fourth joint conference on lexical and computational semantics*. 2015, pp. 117–126.

[63] DeMiguel, Victor, Garlappi, Lorenzo, and Uppal, Raman. "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" In: *The review of Financial studies* 22.5 (2007), pp. 1915–1953.

[64] DeMiguel, Victor et al. "A Generalized Approach to Portfolio Optimization: Improving Performance by Constraining Portfolio Norms". In: *Management Science* 55.5 (2009), pp. 798–812. DOI: 10.1287/mnsc.1080.0986. eprint: https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.1080.0986. URL: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0986.

[65] Di Matteo, Tiziana, Pozzi, Francesca, and Aste, Tomaso. "The use of dynamical networks to detect the hierarchical organization of financial market sectors". In: *The European Physical Journal B* 73.1 (2010), pp. 3–11.

[66] Dias, João. "Sovereign debt crisis in the European Union: A minimum spanning tree approach". In: *Physica A: Statistical Mechanics and its Applications* 391.5 (2012), pp. 2046 –2055. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.11.004. URL: http://www.sciencedirect.com/science/article/pii/S0378437111008399.

[67] Dias, João. "Spanning trees and the Eurozone crisis". In: *Physica A: Statistical Mechanics and its Applications* 392.23 (2013), pp. 5974 –5984. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2013.08.001. URL: http://www.sciencedirect.com/science/article/pii/S0378437113007127.

[68] Dimitrios, Kydros and Vasileios, Oumbailis. "A Network Analysis of the Greek Stock Market". In: *Procedia Economics and Finance* 33 (2015). The Economies of Balkan and Eastern Europe Countries in the Changed World (EBEEC 2015), pp. 340 –349. ISSN: 2212-5671. DOI: https://doi.org/10.1016/S2212-5671(15)01718-9. URL: http://www.sciencedirect.com/science/article/pii/S2212567115017189.

[69] Djauhari, Maman A. "A robust filter in stock networks analysis". In: *Physica A: Statistical Mechanics and its Applications* 391.20 (2012), pp. 5049 –5057. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.05.060. URL: http://www.sciencedirect.com/science/article/pii/S0378437112004323.

[70] Djauhari, Maman Abdurachman and Gan, Siew Lee. "Optimality problem of network topology in stocks market analysis". In: *Physica A: Statistical Mechanics and its Applications* 419 (2015), pp. 108 –114. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.09.060. URL: http://www.sciencedirect.com/science/article/pii/S037843711400836X.

[71] Dose, Christian and Cincotti, Silvano. "Clustering of financial time series with application to index and enhanced index tracking portfolio". In: *Physica A: Statistical Mechanics and its Applications* 355.1 (2005). Market Dynamics and Quantitative Economics, pp. 145 –151. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2005.02.078. URL: http://www.sciencedirect.com/science/article/pii/S0378437105002864.

[72] Drożdż, S et al. "Dynamics of competition between collectivity and noise in the stock market". In: *Physica A: Statistical Mechanics and its Applications* 287.3 (2000), pp. 440 –449. ISSN: 0378-4371. DOI: https://doi.org/10.1016/S0378-4371(00)00383-6. URL: http://www.sciencedirect.com/science/article/pii/S0378437100003836.

[73] Efron, Bradley, Hastie, Trevor, and Tibshirani, Robert. "Discussion: The Dantzig selector: Statistical estimation when p is much larger than n". In: *Ann. Statist.* 35.6 (Dec. 2007), pp. 2358–2364. DOI: 10.1214/009053607000000433. URL: https://doi.org/10.1214/009053607000000433.

[74] Epps, Thomas W. "Comovements in Stock Prices in the Very Short Run". In: *Journal of the American Statistical Association* 74.366 (1979), pp. 291–298. ISSN: 01621459. URL: http://www.jstor.org/stable/2286325.

[75] Erkan, Günes and Radev, Dragomir R. "LexRank: Graph-based Lexical Centrality As Salience in Text Summarization". In: *J. Artif. Int. Res.* 22.1 (Dec. 2004), pp. 457–479. ISSN: 1076-9757. URL: http://dl.acm.org/citation.cfm?id=1622487.1622501.

[76] Esmaeilpour Moghadam, Hadi et al. "Complex networks analysis in Iran stock market: The application of centrality". In: *Physica A: Statistical Mechanics and its Applications* 531 (2019), p. 121800. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2019.121800. URL: http://www.sciencedirect.com/science/article/pii/S0378437119310581.

[77] Estrada, Ernesto. "Quantifying network heterogeneity". In: *Physical Review E* 82.6 (2010), p. 066102.

[78] Eun, Cheol S. and Shim, Sangdal. "International Transmission of Stock Market Movements". In: *Journal of Financial and Quantitative Analysis* 24.2 (1989), 241–256. DOI: 10.2307/2330774.

[79] *Federal Reserve Bank of St. Louis' Financial Crisis Timeline*. URL: https://fraser.stlouisfed.org/timeline/financial-crisis.

[80] Fiedor, Paweł. "Information-theoretic approach to lead-lag effect on financial markets". In: *The European Physical Journal B* 87.8 (2014), p. 168.

[81] Fiedor, Paweł. "Mutual Information-Based Hierarchies on Warsaw Stock Exchange." In: *Acta Physica Polonica, A.* 127 (2015).

[82] Florescu, Corina and Caragea, Cornelia. "Positionrank: An unsupervised approach to keyphrase extraction from scholarly documents". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2017, pp. 1105–1115.

[83] Frank, Eibe et al. "Domain-specific keyphrase extraction". In: *16th International joint conference on artificial intelligence (IJCAI 99)*. Vol. 2. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. 1999, pp. 668–673.

[84] Friedman, Jerome, Hastie, Trevor, and Tibshirani, Rob. "Regularization paths for generalized linear models via coordinate descent". In: *Journal of statistical software* 33.1 (2010), p. 1.

[85] Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. "Applications of the Lasso and grouped Lasso to the estimation of sparse graphical models". In: (Apr. 2010).

[86] Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. "Sparse inverse covariance estimation with the graphical lasso". In: *Biostatistics* 9.3 (2008), pp. 432–441.

[87] Friedman, Jerome, Hastie, Trevor, and Tibshirani, Robert. "Sparse inverse covariance estimation with the graphical LASSO". In: *Biostatistics* 9 (Aug. 2008), pp. 432–41.

[88] Friedman, Jerome et al. "Pathwise coordinate optimization". In: *The annals of applied statistics* 1.2 (2007), pp. 302–332.

[89] Galazka, Marek. "Characteristics of the Polish Stock Market correlations". In: *International Review of Financial Analysis* 20.1 (2011), pp. 1 –5. ISSN: 1057-5219. DOI: https://doi.org/10.1016/j.irfa.2010.11.002. URL: http://www.sciencedirect.com/science/article/pii/S1057521910000827.

[90] Gan, Siew Lee and Djauhari, Maman Abdurachman. "New York Stock Exchange performance: evidence from the forest of multidimensional minimum spanning trees". In: *Journal of Statistical Mechanics: Theory and Experiment* 2015.12 (2015), P12005. DOI: 10.1088/1742-5468/2015/12/p12005. URL: https://doi.org/10.1088%2F1742-5468%2F2015%2F12%2Fp12005.

[91] Gao, Ya-Chun, Zeng, Yong, and Cai, Shi-Min. "Influence network in the Chinese stock market". In: *Journal of Statistical Mechanics: Theory and Experiment* 2015.3 (2015), P03017. DOI: 10.1088/1742-5468/2015/03/p03017. URL: https://doi.org/10.1088%2F1742-5468%2F2015%2F03%2Fp03017.

[92] Gilmore, Claire G., Lucey, Brian M., and Boscia, Marian W. "Comovements in government bond markets: A minimum spanning tree analysis". In: *Physica A: Statistical Mechanics and its Applications* 389.21 (2010), pp. 4875 –4886. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.06.057. URL: http://www.sciencedirect.com/science/article/pii/S0378437110006059.

[93] Gómez, Sergio, Jensen, Pablo, and Arenas, Alex. "Analysis of community structure in networks of correlated data". In: *Physical Review E* 80.1 (2009), p. 016114.

[94] Goto, Shingo and Xu, Yan. "Improving Mean Variance Optimization through Sparse Hedging Restrictions". In: *Journal of Financial and Quantitative Analysis* 50.6 (2015), 1415–1441.

[95] Guo, Xue, Zhang, Hu, and Tian, Tianhai. "Development of stock correlation networks using mutual information and financial big data". In: *PLOS ONE* 13.4 (Apr. 2018), pp. 1–16.

[96] Hagberg, Aric A., Schult, Daniel A., and Swart, Pieter J. "Exploring Network Structure, Dynamics, and Function using NetworkX". In: *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA, 2008, pp. 11 –15. DOI: 10.25080/issn.2575-9752.

[97] Hallac, David et al. "Network Inference via the Time-Varying Graphical Lasso". In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17. Halifax, NS, Canada: ACM, 2017, pp. 205–213. ISBN: 978-1-4503-4887-4.

[98] Han, Juhyun, Kim, Taehwan, and Choi, Joongmin. "Web document clustering by using automatic keyphrase extraction". In: *Web Intelligence and Intelligent Agent Technology Workshops, 2007 IEEE/WIC/ACM International Conferences on*. IEEE. 2007, pp. 56–59.

[99] Hanneman, R. "Introduction to Social Network Methods". In: 2001.

[100] Hartman, David and Hlinka, Jaroslav. "Nonlinearity in stock networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.8 (2018), p. 083127. DOI: 10.1063/1.5023309. eprint: https://doi.org/10.1063/1.5023309. URL: https://doi.org/10.1063/1.5023309.

[101] Hatipoğlu, Veysel Fuat. "Application of a New Quantitative Approach to Stock Markets: Minimum Spanning Tree". In: *Alphanumeric Journal* 5.1 (2017), pp. 163–169.

[102] Heiberger, Raphael H. "Stock network stability in times of crisis". In: *Physica A: Statistical Mechanics and its Applications* 393.C (2014), pp. 376–381. DOI: 10.1016/j.physa.2013.08.0. URL: https://ideas.repec.org/a/eee/phsmap/v393y2014icp376-381.html.

[103] Heimo, Tapio et al. "Spectral methods and cluster structure in correlation-based networks". In: *Physica A: Statistical Mechanics and its Applications* 387.23 (2008), pp. 5930 –5945. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2008.06.028. URL: http://www.sciencedirect.com/science/article/pii/S0378437108005761.

[104] Hlinka, Jaroslav et al. "Small-world bias of correlation networks: From brain to climate". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.3 (2017), p. 035812.

[105] Holmes, Andrew P et al. "Nonparametric analysis of statistic images from functional mapping experiments". In: *Journal of Cerebral Blood Flow & Metabolism* 16.1 (1996), pp. 7–22.

[106] Hsieh, Cho jui et al. "Sparse Inverse Covariance Matrix Estimation Using Quadratic Approximation". In: *Advances in Neural Information Processing Systems 24*. Ed. by Shawe-Taylor, J. et al. Curran Associates, Inc., 2011, pp. 2330–2338. URL: http://papers.nips.cc/paper/4266-sparse-inverse-covariance-matrix-estimation-using-quadratic-approximation.pdf.

[107] Hsieh, Cho-Jui et al. "BIG & QUIC: Sparse Inverse Covariance Estimation for a Million Variables". In: *Advances in Neural Information Processing Systems 26*. Ed. by Burges, C. J. C. et al. Curran Associates, Inc., 2013, pp. 3165–3173. URL: http://papers.nips.cc/paper/4923-big-quic-sparse-inverse-covariance-estimation-for-a-million-variables.pdf.

[108] Huang, Wei-Qiang, Zhuang, Xin-Tian, and Yao, Shuang. "A network analysis of the Chinese stock market". In: *Physica A: Statistical Mechanics and its Applications* 388.14 (2009), pp. 2956 –2964.

[109] Hulth, Anette. "Improved Automatic Keyword Extraction Given More Linguistic Knowledge". In: *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. EMNLP '03. Stroudsburg, PA, USA: Association for Computational Linguistics, 2003, pp. 216–223. DOI: 10.3115/1119355.1119383. URL: https://doi.org/10.3115/1119355.1119383.

[110] Hunter, J. D. "Matplotlib: A 2D Graphics Environment". In: *Computing in Science Engineering* 9.3 (2007), pp. 90–95. ISSN: 1521-9615. DOI: 10.1109/MCSE.2007.55.

[111] Hüttner, Amelie, Mai, Jan-Frederik, and Mineo, Stefano. "Portfolio selection based on graphs: Does it align with Markowitz-optimal portfolios?" In: *Dependence Modeling* 6.1 (2018), pp. 63–87.

[112] Isogai, Takashi. "Building a dynamic correlation network for fat-tailed financial asset returns". In: *Applied Network Science* 1.1 (2016), p. 7. ISSN: 2364-8228. DOI: 10.1007/s41109-016-0008-x. URL: https://doi.org/10.1007/s41109-016-0008-x.

[113] Isogai, Takashi. "Clustering of Japanese stock returns by recursive modularity optimization for efficient portfolio diversification*". In: *Journal of Complex Networks* 2.4 (July 2014), pp. 557–584. ISSN: 2051-1329. DOI: 10.1093/comnet/cnu023. eprint: http://oup.prod.sis.lan/comnet/article-pdf/2/4/557/9130958/cnu023.pdf. URL: https://doi.org/10.1093/comnet/cnu023.

[114] Jagannathan, Ravi and Ma, Tongshu. *Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps*. Working Paper 8922. National Bureau of Economic Research, 2002. DOI: 10.3386/w8922.

[115] Jang, Wooseok, Lee, Junghoon, and Chang, Woojin. "Currency crises and the evolution of foreign exchange market: Evidence from minimum spanning tree". In: *Physica A: Statistical Mechanics and its Applications* 390.4 (2011), pp. 707 –718. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.10.028. URL: http://www.sciencedirect.com/science/article/pii/S0378437110008861.

[116] Ji, Qiang and Fan, Ying. "Evolution of the world crude oil market integration: A graph theory analysis". In: *Energy Economics* 53 (2016). Energy Markets, pp. 90 –100. ISSN: 0140-9883. DOI: https://doi.org/10.1016/j.eneco.2014.12.003. URL: http://www.sciencedirect.com/science/article/pii/S0140988314003193.

[117] Jiang, XF and Zheng, B. "Anti-correlation and subsector structure in financial systems". In: *EPL (Europhysics Letters)* 97.4 (2012), p. 48006.

[118] Jones, Eric, Oliphant, Travis, Peterson, Pearu, et al. *SciPy: Open source scientific tools for Python*. 2001. URL: http://www.scipy.org/.

[119] Jung, Woo-Sung et al. "Characteristics of the Korean stock market correlations". In: *Physica A: Statistical Mechanics and its Applications* 361.1 (2006), pp. 263 –271. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2005.06.081. URL: http://www.sciencedirect.com/science/article/pii/S0378437105007181.

[120] Jung, Woo-Sung et al. "Group dynamics of the Japanese market". In: *Physica A: Statistical Mechanics and its Applications* 387.2 (2008), pp. 537 –542. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2007.09.022. URL: http://www.sciencedirect.com/science/article/pii/S0378437107010114.

[121] Kalyagin, VA et al. "Measures of uncertainty in market network analysis". In: *Physica A: Statistical Mechanics and its Applications* 413 (2014), pp. 59–70.

[122] Kantar, Ersin, Keskin, Mustafa, and Deviren, Bayram. "Analysis of the effects of the global financial crisis on the Turkish economy, using hierarchical methods". In: *Physica A: Statistical Mechanics and its Applications* 391.7 (2012), pp. 2342 –2352. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.12.014. URL: http://www.sciencedirect.com/science/article/pii/S0378437111009113.

[123] Kaya, Hakan. "Eccentricity in asset management". In: *Available at SSRN 2350429* (2013).

[124] Kenett, Dror Y, Shapira, Yoash, and Ben-Jacob, Eshel. "RMT assessments of the market latent information embedded in the stocks' raw, normalized, and partial correlations". In: *Journal of Probability and Statistics* 2009 (2009).

[125] Kenett, Dror Y. et al. "Dependency Network and Node Influence: Application to the Study of Financial Markets". In: *International Journal of Bifurcation and Chaos* 22.07 (2012), p. 1250181. DOI: 10.1142/S0218127412501817. eprint: https://doi.org/10.1142/S0218127412501817. URL: https://doi.org/10.1142/S0218127412501817.

[126] Kenett, Dror Y. et al. "Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market". In: *PLOS ONE* 5.12 (Dec. 2010), pp. 1–14. DOI: 10.1371/journal.pone.0015032.

[127] Kenett, Dror Y et al. "Dynamics of stock market correlations". In: *AUCO Czech Economic Review* 4.3 (2010), pp. 330–341.

[128] Kenett, Dror Y. et al. "Partial correlation analysis: applications for financial markets". In: *Quantitative Finance* 15.4 (2015), pp. 569–578. DOI: 10.1080/14697688.2014.946660. eprint: https://doi.org/10.1080/14697688.2014.946660. URL: https://doi.org/10.1080/14697688.2014.946660.

[129] Keskin, Mustafa, Deviren, Bayram, and Kocakaplan, Yusuf. "Topology of the correlation networks among major currencies using hierarchical structure methods". In: *Physica A: Statistical Mechanics and its Applications* 390.4 (2011), pp. 719 –730. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.10.041. URL: http://www.sciencedirect.com/science/article/pii/S0378437110009209.

[130] Khoojine, Arash Sioofy and Han, Dong. "Stock price network autoregressive model with application to stock market turbulence". In: *The European Physical Journal B* 93.7 (2020), pp. 1–15.

[131] Kim, Hyun-Joo et al. "Weighted Scale-Free Network in Financial Correlations". In: *Journal of the Physical Society of Japan* 71.9 (2002), pp. 2133–2136. DOI: 10.1143/JPSJ.71.2133. eprint: https://doi.org/10.1143/JPSJ.71.2133. URL: https://doi.org/10.1143/JPSJ.71.2133.

[132] Kim, Su Nam et al. "Automatic keyphrase extraction from scientific articles". In: *Language Resources and Evaluation* 47.3 (2013), pp. 723–742. ISSN: 1574-0218. DOI: 10.1007/s10579-012-9210-3. URL: https://doi.org/10.1007/s10579-012-9210-3.

[133] Kocakaplan, Yusuf et al. "Correlations, hierarchies and networks of the world's automotive companies". In: *Physica A: Statistical Mechanics and its Applications* 392.12 (2013), pp. 2736 –2774. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2013.02.013. URL: http://www.sciencedirect.com/science/article/pii/S0378437113001714.

[134] Kocheturov, Anton, Batsyn, Mikhail, and Pardalos, Panos M. "Dynamics of cluster structures in a financial market network". In: *Physica A: Statistical Mechanics and its Applications* 413 (2014), pp. 523 –533. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.06.077. URL: http://www.sciencedirect.com/science/article/pii/S0378437114005585.

[135] Kritzman, Mark, Page, Sébastien, and Turkington, David. "In Defense of Optimization: The Fallacy of 1/N". In: *Financial Analysts Journal* 66.2 (2010), pp. 31–39. DOI: 10.2469/faj.v66.n2.6. eprint: https://doi.org/10.2469/faj.v66.n2.6. URL: https://doi.org/10.2469/faj.v66.n2.6.

[136] Kulkarni, Varsha and Deo, Nivedita. "Correlation and volatility in an Indian stock market: A random matrix approach". In: *The European Physical Journal B* 60.1 (2007), pp. 101–109.

[137] Kullmann, L, Kertész, Janos, and Kaski, K. "Time-dependent cross-correlations between different stock returns: A directed network of influence". In: *Physical Review E* 66.2 (2002), p. 026125.

[138] Kwapien, Jaroslaw, Gworek, Sylwia, and Drozdz, Stanislaw. "Structure and evolution of the foreign exchange networks". In: *arXiv preprint arXiv:0901.4793* (2009).

[139] Laloux, Laurent et al. "Random matrix theory and financial correlations". In: *International Journal of Theoretical and Applied Finance* 3.03 (2000), pp. 391–397.

[140] Landauer, Thomas K, Foltz, Peter W, and Laham, Darrell. "An introduction to latent semantic analysis". In: *Discourse processes* 25.2-3 (1998), pp. 259–284.

[141] Langfelder, Peter and Horvath, Steve. "WGCNA: an R package for weighted correlation network analysis". In: *BMC Bioinformatics* 9.1 (2008), p. 559. ISSN: 1471-2105. DOI: 10.1186/1471-2105-9-559. URL: https://doi.org/10.1186/1471-2105-9-559.

[142] Ledoit, Olivier and Wolf, Michael. "A well-conditioned estimator for large-dimensional covariance matrices". In: *Journal of Multivariate Analysis* 88.2 (2004), pp. 365–411.

[143] Ledoit, Olivier and Wolf, Michael. "Honey, I Shrunk the Sample Covariance Matrix". In: *The Journal of Portfolio Management* 30.4 (2004), pp. 110–119.

[144] Lee, Gan Siew and Djauhari, Maman A. "Multidimensional stock network analysis: An Escoufier's RV coefficient approach". In: *AIP Conference Proceedings* 1557.1 (2013), pp. 550–555. DOI: 10.1063/1.4823975. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.4823975. URL: https://aip.scitation.org/doi/abs/10.1063/1.4823975.

[145] Lee, Junghoon, Youn, Janghyuk, and Chang, Woojin. "Intraday volatility and network topological properties in the Korean stock market". In: *Physica A: Statistical Mechanics and its Applications* 391.4 (2012), pp. 1354 –1360. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.09.016. URL: http://www.sciencedirect.com/science/article/pii/S0378437111007321.

[146] L'Huillier, G. et al. "Latent semantic analysis and keyword extraction for phishing classification". In: *2010 IEEE International Conference on Intelligence and Security Informatics*. 2010, pp. 129–131. DOI: 10.1109/ISI.2010.5484762.

[147] Li, Bentian and Pi, Dechang. "Analysis of global stock index data during crisis period via complex network approach". In: *PLOS ONE* 13.7 (July 2018), pp. 1–16. DOI: 10.1371/journal.pone.0200600. URL: https://doi.org/10.1371/journal.pone.0200600.

[148] Li, Chang-Jin and Han, Hui-Jian. "Keyword Extraction Algorithm Based on Principal Component Analysis". In: *Intelligent Computing and Information Science*. Ed. by Chen, Ran. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 503–508. ISBN: 978-3-642-18134-4.

[149] Li, Linyue, Willett, Thomas D, and Zhang, Nan. "The effects of the global financial crisis on China's financial market and macroeconomy". In: *Economics Research International* 2012 (2012).

[150] Li, Yan et al. "Portfolio optimization based on network topology". In: *Physica A: Statistical Mechanics and its Applications* 515 (2019), pp. 671 –681. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2018.10.014. URL: http://www.sciencedirect.com/science/article/pii/S0378437118313529.

[151] Liu, Han, Lafferty, John, and Wasserman, Larry. "The nonparanormal: Semiparametric estimation of high dimensional undirected graphs". In: *Journal of Machine Learning Research* 10.Oct (2009), pp. 2295–2328.

[152] Liu, Han and Wang, Lie. "TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models". In: *Electron. J. Statist.* 11.1 (2017), pp. 241–294. DOI: 10.1214/16-EJS1195. URL: https://doi.org/10.1214/16-EJS1195.

[153] Liu, Weidong and Luo, Xi. "Fast and adaptive sparse precision matrix estimation in high dimensions". In: *Journal of Multivariate Analysis* 135 (2015), pp. 153 –162. ISSN: 0047-259X. DOI: https://doi.org/10.1016/j.jmva.2014.11.005. URL: http://www.sciencedirect.com/science/article/pii/S0047259X14002607.

[154] Lo, Andrew W and MacKinlay, A. Craig. *Stock Market Prices Do Not Follow Random Walks: Evidence From a Simple Specification Test*. Working Paper 2168. National Bureau of Economic Research, 1987. DOI: 10.3386/w2168. URL: http://www.nber.org/papers/w2168.

[155] Lorrain, Francois and White, Harrison C. "Structural equivalence of individuals in social networks". In: *The Journal of mathematical sociology* 1.1 (1971), pp. 49–80.

[156] Lü, Linyuan and Zhou, Tao. "Link prediction in complex networks: A survey". In: *Physica A: statistical mechanics and its applications* 390.6 (2011), pp. 1150–1170.

[157] MacMahon, Mel and Garlaschelli, Diego. "Community Detection for Correlation Matrices". In: *Phys. Rev. X* 5 (2 2015), p. 021006. DOI: 10.1103/PhysRevX.5.021006. URL: https://link.aps.org/doi/10.1103/PhysRevX.5.021006.

[158] Mai, Yong et al. "Currency co-movement and network correlation structure of foreign exchange market". In: *Physica A: Statistical Mechanics and its Applications* 492 (2018), pp. 65 –74. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2017.09.068. URL: http://www.sciencedirect.com/science/article/pii/S0378437117309603.

[159] Majapa, Mohamed and Gossel, Sean Joss. "Topology of the South African stock market network across the 2008 financial crisis". In: *Physica A: Statistical Mechanics and its Applications* 445 (2016), pp. 35 –47. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.10.108. URL: http://www.sciencedirect.com/science/article/pii/S0378437115009784.

[160] Mantegna, R. N. "Hierarchical structure in financial markets". In: *The European Physical Journal B - Condensed Matter and Complex Systems* 11.1 (1999), pp. 193–197.

[161] Markowitz, H. M. "Portfolio selection". In: *The Journal of Finance* 7.60 (1952), pp. 77–91.

[162] Marti, Gautier et al. "A review of two decades of correlations, hierarchies, networks and clustering in financial markets". In: *arXiv preprint arXiv:1703.00485* (2017).

[163] Massara, Guido Previde, Di Matteo, Tiziana, and Aste, Tomaso. "Network filtering for big data: Triangulated maximally filtered graph". In: *Journal of complex Networks* 5.2 (2016), pp. 161–178.

[164] Matia, K et al. "Scale-dependent price fluctuations for the Indian stock market". In: *Europhysics Letters (EPL)* 66.6 (2004), pp. 909–914. DOI: 10.1209/epl/i2003-10267-y. URL: https://doi.org/10.1209%2Fepl%2Fi2003-10267-y.

[165] MATSUO, Y. and ISHIZUKA, M. "KEYWORD EXTRACTION FROM A SINGLE DOCUMENT USING WORD CO-OCCURRENCE STATISTICAL INFORMATION". In: *International Journal on Artificial Intelligence Tools* 13.01 (2004), pp. 157–169. DOI: 10.1142/S0218213004001466. eprint: https://doi.org/10.1142/S0218213004001466. URL: https://doi.org/10.1142/S0218213004001466.

[166] Mazumder, Rahul and Hastie, Trevor. "Exact Covariance Thresholding into Connected Components for Large-scale Graphical Lasso". In: *J. Mach. Learn. Res.* 13.1 (Mar. 2012), pp. 781–794. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=2503308.2188412.

[167] Mazumder, Rahul and Hastie, Trevor. "The graphical lasso: New insights and alternatives". In: *Electron. J. Statist.* 6 (2012), pp. 2125–2149.

[168] McDonald, Mark et al. "Detecting a currency's dominance or dependence using foreign exchange network trees". In: *Phys. Rev. E* 72 (4 2005), p. 046106. DOI: 10.1103/PhysRevE.72.046106. URL: https://link.aps.org/doi/10.1103/PhysRevE.72.046106.

[169] McKinney, Wes. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. 2010, pp. 51 –56.

[170] Meinshausen, Nicolai and Buhlmann, Peter. "High-dimensional graphs and variable selection with the Lasso". In: *The Annals of Statistics* 34.3 (June 2006), pp. 1436–1462. DOI: 10.1214/009053606000000281.

[171] Meng, Hao et al. "Systemic risk and spatiotemporal dynamics of the US housing market". In: *Scientific reports* 4 (2014), p. 3655.

[172] Miccichè, Salvatore et al. "Degree stability of a minimum spanning tree of price return and volatility". In: *Physica A: Statistical Mechanics and its Applications* 324.1 (2003). Proceedings of the International Econophysics Conference, pp. 66 –73. ISSN: 0378-4371. DOI: https://doi.org/10.1016/S0378-4371(03)00002-5. URL: http://www.sciencedirect.com/science/article/pii/S0378437103000025.

[173] Mihalcea, Rada and Tarau, Paul. "TextRank: Bringing Order into Text". In: *Proceedings of EMNLP 2004*. Barcelona, Spain: Association for Computational Linguistics, July 2004, pp. 404–411. URL: https://www.aclweb.org/anthology/W04-3252.

[174] Millington, Tristan and Niranjan, Mahesan. "Partial correlation financial networks". In: *Applied Network Science* 5.1 (2020), p. 11.

[175] Millington, Tristan and Niranjan, Mahesan. "Robust Portfolio Risk Minimization Using the Graphical Lasso". In: *Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II*. Cham: Springer International Publishing, 2017, pp. 863–872. ISBN: 978-3-319-70096-0.

[176] Mizuno, Takayuki, Takayasu, Hideki, and Takayasu, Misako. "Correlation networks among currencies". In: *Physica A: Statistical Mechanics and its Applications* 364 (2006), pp. 336–342.

[177] Musciotto, F. et al. "Bootstrap validation of links of a minimum spanning tree". In: *Physica A: Statistical Mechanics and its Applications* 512 (2018), pp. 1032 –1043. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2018.08.020. URL: http://www.sciencedirect.com/science/article/pii/S0378437118309695.

[178] Musmeci, Nicoló, Aste, Tomaso, and Di Matteo, Tiziana. "Interplay between past market correlation structure changes and future volatility outbursts". In: *Scientific reports* 6 (2016), p. 36320.

[179] Musmeci, Nicoló, Aste, Tomaso, and Di Matteo, Tiziana. "Risk diversification: a study of persistence with a filtered correlation-network approach". In: *arXiv preprint arXiv:1410.5621* (2014).

[180] Musmeci, Nicoló et al. "The multiplex dependency structure of financial markets". In: *Complexity* 2017 (2017).

[181] Musmeci, Nicoló, Aste, Tomaso, and Di Matteo, T. "Relation between Financial Market Structure and the Real Economy: Comparison between Clustering Methods". In: *PLOS ONE* 10.3 (Mar. 2015), pp. 1–24. DOI: 10.1371/journal.pone.0116201. URL: https://doi.org/10.1371/journal.pone.0116201.

[182] Namaki, A. et al. "Network analysis of a financial market based on genuine correlation and threshold method". In: *Physica A: Statistical Mechanics and its Applications* 390.21 (2011), pp. 3835 –3841. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.06.033. URL: http://www.sciencedirect.com/science/article/pii/S0378437111004808.

[183] Naylor, Michael J., Rose, Lawrence C., and Moyle, Brendan J. "Topology of foreign exchange markets using hierarchical structure methods". In: *Physica A: Statistical Mechanics and its Applications* 382.1 (2007). Applications of Physics in Financial Analysis, pp. 199 –208. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2007.02.019. URL: http://www.sciencedirect.com/science/article/pii/S0378437107001513.

[184] Newman, M. E. J. "Modularity and community structure in networks". In: *Proceedings of the National Academy of Sciences* 103.23 (2006), pp. 8577–8582. ISSN: 0027-8424. DOI: 10.1073/pnas.0601602103. eprint: https://www.pnas.org/content/103/23/8577.full.pdf. URL: https://www.pnas.org/content/103/23/8577.

[185] Newman, Mark. *Networks: an introduction*. Oxford university press, 2010. DOI: 10.1017/CBO9780511761942.

[186] Nguyen, Thuy Dung and Kan, Min-Yen. "Keyphrase Extraction in Scientific Publications". In: *Proceedings of the 10th International Conference on Asian Digital Libraries: Looking Back 10 Years and Forging New Frontiers*. ICADL'07. Hanoi, Vietnam: Springer-Verlag, 2007, pp. 317–326. ISBN: 3-540-77093-3, 978-3-540-77093-0. URL: http://dl.acm.org/citation.cfm?id=1780653.1780707.

[187] Nie, Chun-Xiao and Song, Fu-Tie. "Constructing financial network based on PMFG and threshold method". In: *Physica A: Statistical Mechanics and its Applications* 495 (2018), pp. 104 –113. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2017.12.037. URL: http://www.sciencedirect.com/science/article/pii/S037843711731292X.

[188] Nie, Chun-xiao and Song, Fu-tie. "Relationship between Entropy and Dimension of Financial Correlation-Based Network". In: *Entropy* 20.3 (2018). ISSN: 1099-4300. DOI: 10.3390/e20030177. URL: http://dx.doi.org/10.3390/e20030177.

[189] Nie, Chun-Xiao, Song, Fu-Tie, and Li, Sai-Ping. "Rényi indices of financial minimum spanning trees". In: *Physica A: Statistical Mechanics and its Applications* 444 (2016), pp. 883 –889. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.10.087. URL: http://www.sciencedirect.com/science/article/pii/S0378437115009462.

[190] Nobi, Ashadun and Lee, Jae Woo. "Systemic risk and hierarchical transitions of financial networks". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.6 (2017), p. 063107. DOI: 10.1063/1.4978925. eprint: https://doi.org/10.1063/1.4978925. URL: https://doi.org/10.1063/1.4978925.

[191] Nobi, Ashadun et al. "Effects of global financial crisis on network structure in a local stock market". In: *Physica A: Statistical Mechanics and its Applications* 407 (2014), pp. 135 –143. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.03.083. URL: http://www.sciencedirect.com/science/article/pii/S0378437114002945.

[192] Oh, Gabjin, Kim, Seunghwan, and Eom, Cheoljun. "Market efficiency in foreign exchange markets". In: *Physica A: Statistical Mechanics and its Applications* 382.1 (2007). Applications of Physics in Financial Analysis, pp. 209 –212. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2007.02.032. URL: http://www.sciencedirect.com/science/article/pii/S0378437107001525.

[193] Oliphant, Travis E. *A guide to NumPy*. Vol. 1. 2006.

[194] Onnela, J-P, Kaski, Kimmo, and Kertész, Janos. "Clustering and information in correlation based financial networks". In: *The European Physical Journal B* 38.2 (2004), pp. 353–362.

[195] Onnela, J.-P. et al. "Dynamic asset trees and Black Monday". In: *Physica A: Statistical Mechanics and its Applications* 324.1 (2003). Proceedings of the International Econophysics Conference, pp. 247 –252.

[196] Onnela, J-P et al. "Dynamics of market correlations: Taxonomy and portfolio analysis". In: *Physical Review E* 68.5 (2003), p. 056110.

[197] Ontañón, Santiago. "An overview of distance and similarity functions for structured data". In: *Artificial Intelligence Review* (2020), pp. 1–43.

[198] ORTEGA, GUILLERMO J. and MATESANZ, DAVID. "CROSS-COUNTRY HIERARCHICAL STRUCTURE AND CURRENCY CRISES". In: *International Journal of Modern Physics C* 17.03 (2006), pp. 333–341. DOI: 10.1142/S012918310600856X. eprint: https://doi.org/10.1142/S012918310600856X. URL: https://doi.org/10.1142/S012918310600856X.

[199] Papenbrock, Jochen and Schwendner, Peter. "Handling risk-on/risk-off dynamics with correlation regimes and correlation networks". In: *Financial Markets and Portfolio Management* 29.2 (2015), pp. 125–147. ISSN: 2373-8529. DOI: 10.1007/s11408-015-0248-2. URL: https://doi.org/10.1007/s11408-015-0248-2.

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

[201] Pedregosa, Fabian et al. "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12.Oct (2011), pp. 2825–2830.

[202] Peng, Jie et al. "Partial Correlation Estimation by Joint Sparse Regression Models". In: *Journal of the American Statistical Association* 104.486 (2009). PMID: 19881892, pp. 735–746. DOI: 10.1198/jasa.2009.0126. eprint: https://doi.org/10.1198/jasa.2009.0126. URL: https://doi.org/10.1198/jasa.2009.0126.

[203] Peralta, Gustavo. "Network-based measures as leading indicators of market instability: the case of the Spanish stock market". In: *Journal of Network Theory in Finance* (2015).

[204] Peralta, Gustavo and Zareei, Abalfazl. "A network approach to portfolio selection". In: *Journal of Empirical Finance* 38 (2016), pp. 157 –180.

[205] Piccardi, Carlo, Calatroni, Lisa, and Bertoni, Fabio. "Clustering Financial Time Series By Network Community Analysis". In: *International Journal of Modern Physics C* 22.01 (2011), pp. 35–50. DOI: 10.1142/S012918311101604X. eprint: https://doi.org/10.1142/S012918311101604X. URL: https://doi.org/10.1142/S012918311101604X.

[206] Plerou, Vasiliki et al. "Random matrix approach to cross correlations in financial data". In: *Physical Review E* 65.6 (2002), p. 066126.

[207] Podobnik, Boris and Stanley, H Eugene. "Detrended cross-correlation analysis: a new method for analyzing two nonstationary time series". In: *Physical review letters* 100.8 (2008), p. 084102.

[208] Potters, Marc, Bouchaud, Jean-Philippe, and Laloux, Laurent. "Financial applications of random matrix theory: Old laces and new pieces". In: *arXiv preprint physics/0507111* (2005).

[209] Pozzi, FDMT, Di Matteo, Tiziana, and Aste, Tomaso. "Exponential smoothing weighted correlations". In: *The European Physical Journal B* 85.6 (2012), p. 175.

[210] Pozzi, Francesco, Di Matteo, Tiziana, and Aste, Tomaso. "Spread of risk across financial markets: better to invest in the peripheries". In: *Scientific reports* 3 (2013), p. 1665.

[211] Preis, Tobias et al. "Quantifying the Behavior of Stock Correlations Under Market Stress". In: *Scientific Reports* 2 (2012). Article, 752 EP –.

[212] Pudota, Nirmala et al. "Automatic Keyphrase Extraction and Ontology Mining for Content-based Tag Recommendation". In: *Int. J. Intell. Syst.* 25.12 (Dec. 2010), pp. 1158–1186. ISSN: 0884-8173. DOI: 10.1002/int.v25:12. URL: http://dx.doi.org/10.1002/int.v25:12.

[213] Qiao, Haishu, Li, Ying, and Xia, Yue. "Analysis of linkage effects among currency networks using REER data". In: *Discrete Dynamics in Nature and Society* 2015 (2015).

[214] Qiu, Huitong et al. "Robust Portfolio Optimization". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*. NIPS'15. Montreal, Canada: MIT Press, 2015, pp. 46–54.

[215] Rand, William M. "Objective Criteria for the Evaluation of Clustering Methods". In: *Journal of the American Statistical Association* 66.336 (1971), pp. 846–850. DOI: 10.1080/01621459.1971.10482356. eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1971.10482356. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1971.10482356.

[216] Robert, P. and Escoufier, Y. "A Unifying Tool for Linear Multivariate Statistical Methods: The RV- Coefficient". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 25.3 (1976), pp. 257–265. ISSN: 00359254, 14679876. URL: http://www.jstor.org/stable/2347233.

[217] Rosa, Maria J. et al. "Sparse network-based models for patient classification using fMRI". In: *NeuroImage* 105 (2015), pp. 493 –506. DOI: 10.1016/j.neuroimage.2014.11.021.

[218] Rose, Stuart et al. "Automatic keyword extraction from individual documents". In: *Text mining: applications and theory* (2010), pp. 1–20.

[219] Ross, Gordon J. "Dynamic multifactor clustering of financial networks". In: *Physical Review E* 89.2 (2014), p. 022809.

[220] Rosvall, Martin and Bergstrom, Carl T. "Maps of information flow reveal community structure in complex networks". In: *arXiv preprint physics.soc-ph/0707.0609* (2007).

[221] Rubin, Danielle N., Bassett, Danielle S., and Ready, Robert. "Uncovering dynamic stock return correlations with multilayer network analysis". In: *Applied Network Science* 4.1 (2019), p. 31. ISSN: 2364-8228. DOI: 10.1007/s41109-019-0132-5. URL: https://doi.org/10.1007/s41109-019-0132-5.

[222] Saeedian, M. et al. "Emergence of world-stock-market network". In: *Physica A: Statistical Mechanics and its Applications* 526 (2019), p. 120792. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2019.04.028. URL: http://www.sciencedirect.com/science/article/pii/S0378437119303942.

[223] Sandoval, Leonidas and Franca, Italo De Paula. "Correlation of financial markets in times of crisis". In: *Physica A: Statistical Mechanics and its Applications* 391.1 (2012), pp. 187 –208. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2011.07.023. URL: http://www.sciencedirect.com/science/article/pii/S037843711100570X.

[224] Schäfer, Juliane and Strimmer, Korbinian. "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics". In: *Statistical applications in genetics and molecular biology* 4.1 (2005).

[225] Schäfer, Juliane and Strimmer, Korbinian. "An empirical Bayes approach to inferring large-scale gene association networks". In: *Bioinformatics* 21.6 (2004), pp. 754–764.

[226] Schwarz, Gideon. "Estimating the Dimension of a Model". In: *Ann. Statist.* 6.2 (Mar. 1978), pp. 461–464. DOI: 10.1214/aos/1176344136. URL: https://doi.org/10.1214/aos/1176344136.

[227] Seabold, Skipper and Perktold, Josef. "Statsmodels: Econometric and statistical modeling with python". In: *9th Python in Science Conference*. 2010.

[228] Shanker, O. "Defining dimension of a complex network". In: *Modern Physics Letters B* 21.06 (2007), pp. 321–326.

[229] Shannon, Paul et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome research* 13.11 (2003), pp. 2498–2504.

[230] Shapira, Y., Kenett, D. Y., and Ben-Jacob, E. "The Index cohesive effect on stock market correlations". In: *The European Physical Journal B* 72.4 (2009), p. 657. ISSN: 1434-6036. DOI: 10.1140/epjb/e2009-00384-y. URL: https://doi.org/10.1140/epjb/e2009-00384-y.

[231] Sharma, Charu and Habib, Amber. "Mutual information based stock networks and portfolio selection for intraday traders using high frequency data: An Indian market case study". In: *PLOS ONE* 14.8 (Aug. 2019), pp. 1–19. DOI: 10.1371/journal.pone.0221910. URL: https://doi.org/10.1371/journal.pone.0221910.

[232] Shen, J. and Zheng, B. "Cross-correlation in financial dynamics". In: *EPL (Europhysics Letters)* 86.4 (2009), p. 48005. DOI: 10.1209/0295-5075/86/48005. URL: https://doi.org/10.1209%2F0295-5075%2F86%2F48005.

[233] Sheppard, Kevin et al. "bashtage/arch: Release 4.8.1". Version 4.8.1. In: (Mar. 2019). DOI: 10.5281/zenodo.2613877. URL: https://doi.org/10.5281/zenodo.2613877.

[234] Shirokikh, Oleg et al. "Computational study of the US stock market evolution: a rank correlation-based network model". In: *Computational Management Science* 10.2 (2013), pp. 81–103. DOI: 10.1007/s10287-012-0160-4. URL: https://ideas.repec.org/a/spr/comgts/v10y2013i2p81-103.html.

[235] Sieczka, Paweł and Hołyst, Janusz A. "Correlations in commodity markets". In: *Physica A: Statistical Mechanics and its Applications* 388.8 (2009), pp. 1621–1630.

[236] Sinha, Sitabhra and Pan, Raj Kumar. *Uncovering the Internal Structure of the Indian Financial Market: Cross-correlation behavior in the NSE*. Papers 0704.2115. arXiv.org, Apr. 2007. URL: https://ideas.repec.org/p/arx/papers/0704.2115.html.

[237] Sojoudi, Somayeh. "Equivalence of Graphical Lasso and Thresholding for Sparse Graphs". In: *Journal of Machine Learning Research* 17.115 (2016), pp. 1–21. URL: http://jmlr.org/papers/v17/16-013.html.

[238] Song, Dong-Ming et al. "Evolution of worldwide stock markets, correlation structure, and correlation-based graphs". In: *Phys. Rev. E* 84 (2 2011), p. 026108. DOI: 10.1103/PhysRevE.84.026108. URL: https://link.aps.org/doi/10.1103/PhysRevE.84.026108.

[239] Song, Jung Yoon, Chang, Woojin, and Song, Jae Wook. "Cluster analysis on the structure of the cryptocurrency market via Bitcoin–Ethereum filtering". In: *Physica A: Statistical Mechanics and its Applications* 527 (2019), p. 121339. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2019.121339. URL: http://www.sciencedirect.com/science/article/pii/S0378437119304893.

[240] Song, Won-Min, Aste, Tomaso, and Di Matteo, Tiziana. "Correlation-based biological networks". In: *Complex Systems II*. Vol. 6802. International Society for Optics and Photonics. 2008, p. 680212.

[241] Song, Won-Min, Di Matteo, T., and Aste, Tomaso. "Hierarchical Information Clustering by Means of Topologically Embedded Graphs". In: *PLOS ONE* 7.3 (Mar. 2012), pp. 1–14. DOI: 10.1371/journal.pone.0031929. URL: https://doi.org/10.1371/journal.pone.0031929.

[242] Song, Won-Min and Zhang, Bin. "Multiscale embedded gene co-expression network analysis". In: *PLoS computational biology* 11.11 (2015).

[243] Souto, Pedro C. et al. "Capturing Financial Volatility Through Simple Network Measures". In: *Complex Networks and Their Applications VII*. Ed. by Aiello, Luca Maria et al. Cham: Springer International Publishing, 2019, pp. 534–546. ISBN: 978-3-030-05414-4.

[244] Sporns, Olaf. "Contributions and challenges for network models in cognitive neuroscience". In: *Nature neuroscience* 17.5 (2014), p. 652.

[245] Stosic, Darko et al. "Collective behavior of cryptocurrency price changes". In: *Physica A: Statistical Mechanics and its Applications* 507 (2018), pp. 499–509.

[246] Sun, Tingni and Zhang, Cun-Hui. "Scaled sparse linear regression". In: *Biometrika* 99.4 (Sept. 2012), pp. 879–898. ISSN: 0006-3444. DOI: 10.1093/biomet/ass043. eprint: http://oup.prod.sis.lan/biomet/article-pdf/99/4/879/627955/ass043.pdf. URL: https://doi.org/10.1093/biomet/ass043.

[247] Tabak, Benjamin M., Serra, Thiago R., and Cajueiro, Daniel O. "Topological properties of stock market networks: The case of Brazil". In: *Physica A: Statistical Mechanics and its Applications* 389.16 (2010), pp. 3240 –3249. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2010.04.002. URL: http://www.sciencedirect.com/science/article/pii/S0378437110002992.

[248] Tabak, BM, Serra, TR, and Cajueiro, DO. "Topological properties of commodities networks". In: *The European Physical Journal B* 74.2 (2010), pp. 243–249.

[249] Takeda, Akiko et al. "Simultaneous pursuit of out-of-sample performance and sparsity in index tracking portfolios". In: *Computational Management Science* 10.1 (2013), pp. 21–49.

[250] Tan, Kean Ming et al. "Learning graphical models with hubs". In: *The Journal of Machine Learning Research* 15.1 (2014), pp. 3297–3331.

[251] Taylor, Stephen J. *Asset price dynamics, volatility, and prediction*. Princeton university press, 2011.

[252] Teh, Boon Kin et al. "The Chinese Correction of February 2007: How financial hierarchies change in a market crash". In: *Physica A: Statistical Mechanics and its Applications* 424 (2015), pp. 225 –241. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.01.024. URL: http://www.sciencedirect.com/science/article/pii/S0378437115000266.

[253] Tewarie, P. et al. "Functional brain network analysis using minimum spanning trees in Multiple Sclerosis: An MEG source-space study". In: *NeuroImage* 88 (2014), pp. 308 –318. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2013.10.022. URL: http://www.sciencedirect.com/science/article/pii/S1053811913010458.

[254] Tewarie, P. et al. "The minimum spanning tree: An unbiased method for brain network analysis". In: *NeuroImage* 104 (2015), pp. 177 –188. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2014.10.015. URL: http://www.sciencedirect.com/science/article/pii/S1053811914008398.

[255] Tibshirani, Robert. "Regression Shrinkage and Selection via the Lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1 (1996), pp. 267–288.

[256] Tola, Vincenzo et al. "Cluster analysis for portfolio optimization". In: *Journal of Economic Dynamics and Control* 32.1 (2008). Applications of statistical physics in economics and finance, pp. 235 –258. ISSN: 0165-1889. DOI: https://doi.org/10.1016/j.jedc.2007.01.034. URL: http://www.sciencedirect.com/science/article/pii/S0165188907000462.

[257] Torri, Gabriele, Giacometti, Rosella, and Paterlini, Sandra. "Sparse precision matrices for minimum variance portfolios". In: *Computational Management Science* 16.3 (2019), pp. 375–400. ISSN: 1619-6988. DOI: 10.1007/s10287-019-00344-6. URL: https://doi.org/10.1007/s10287-019-00344-6.

[258] Treviño Aguilar, Erick. "The interdependency structure in the Mexican stock exchange: A network approach". In: *arXiv* (2020), arXiv–2004.

[259] Trusina, Ala et al. "Hierarchy Measures in Complex Networks". In: *Phys. Rev. Lett.* 92 (17 2004), p. 178702. DOI: 10.1103/PhysRevLett.92.178702. URL: https://link.aps.org/doi/10.1103/PhysRevLett.92.178702.

[260] Tumminello, M. et al. "A tool for filtering information in complex systems". In: *Proceedings of the National Academy of Sciences* 102.30 (2005), pp. 10421–10426.

[261] Tumminello, M. et al. "Correlation based networks of equity returns sampled at different time horizons". In: *The European Physical Journal B* 55.2 (2007), pp. 209–217. ISSN: 1434-6036. DOI: 10.1140/epjb/e2006-00414-4. URL: https://doi.org/10.1140/epjb/e2006-00414-4.

[262] Tumminello, Michele, Lillo, Fabrizio, and Mantegna, Rosario N. "Correlation, hierarchies, and networks in financial markets". In: *Journal of Economic Behavior & Organization* 75.1 (2010). Transdisciplinary Perspectives on Economic Complexity, pp. 40 –58. ISSN: 0167-2681. DOI: https://doi.org/10.1016/j.jebo.2010.01.004. URL: http://www.sciencedirect.com/science/article/pii/S0167268110000077.

[263] Tumminello, Michele et al. "Spanning trees and bootstrap reliability estimation in correlation-based networks". In: *International Journal of Bifurcation and Chaos* 17.07 (2007), pp. 2319–2329.

[264] Turiel, Jeremy D. and Aste, Tomaso. "Sector Neutral Portfolios: Long Memory Motifs Persistence in Market Structure Dynamics". In: *Complex Networks and Their Applications VIII*. Ed. by Cherifi, Hocine et al. Cham: Springer International Publishing, 2020, pp. 573–585. ISBN: 978-3-030-36683-4.

[265] Turney, Peter D. "Learning Algorithms for Keyphrase Extraction". In: *Information Retrieval* 2.4 (2000), pp. 303–336. ISSN: 1573-7659. DOI: 10.1023/A:1009976227802. URL: https://doi.org/10.1023/A:1009976227802.

[266] Uechi, Lisa et al. "Sector dominance ratio analysis of financial markets". In: *Physica A: Statistical Mechanics and its Applications* 421 (2015), pp. 488 –509. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2014.11.055. URL: http://www.sciencedirect.com/science/article/pii/S0378437114010267.

[267] Vandewalle, N, Brisbois, F, Tordoir, X, et al. "Non-random topology of stock markets". In: *Quantitative Finance* 1.3 (2001), pp. 372–374.

[268] Vizgunov, A. et al. "Network approach for the Russian stock market". In: *Computational Management Science* 11.1 (2014), pp. 45–55.

[269] Vodenska, Irena et al. "Community Analysis of Global Financial Markets". In: *Risks* 4.2 (2016), pp. 1–15. URL: https://ideas.repec.org/a/gam/jrisks/v4y2016i2p13-d70032.html.

[270] Vértes, Petra et al. "Topological Isomorphisms of Human Brain and Financial Market Networks". In: *Frontiers in Systems Neuroscience* 5 (2011), p. 75. ISSN: 1662-5137. DOI: 10.3389/fnsys.2011.00075. URL: https://www.frontiersin.org/article/10.3389/fnsys.2011.00075.

[271] Wainwright, M. J. "Sharp Thresholds for High-Dimensional and Noisy Sparsity Recovery Using ℓ_1 -Constrained Quadratic Programming (Lasso)". In: *IEEE Transactions on Information Theory* 55.5 (2009), pp. 2183–2202. ISSN: 0018-9448. DOI: 10.1109/TIT.2009.2016018.

[272] Wan, Xiaojun and Xiao, Jianguo. "Single Document Keyphrase Extraction Using Neighborhood Knowledge." In: *AAAI*. Vol. 8. 2008, pp. 855–860.

[273] Wang, Duan et al. "Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices". In: *Phys. Rev. E* 83 (4 2011), p. 046121. DOI: 10.1103/PhysRevE.83.046121. URL: https://link.aps.org/doi/10.1103/PhysRevE.83.046121.

[274] Wang, Gang-Jin and Xie, Chi. "Correlation structure and dynamics of international real estate securities markets: A network perspective". In: *Physica A: Statistical Mechanics and its Applications* 424 (2015), pp. 176 –193. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.01.025. URL: http://www.sciencedirect.com/science/article/pii/S0378437115000278.

[275] Wang, Gang-Jin and Xie, Chi. "Cross-correlations between Renminbi and four major currencies in the Renminbi currency basket". In: *Physica A: Statistical Mechanics and its Applications* 392.6 (2013), pp. 1418 –1428. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2012.11.035. URL: http://www.sciencedirect.com/science/article/pii/S037843711201000X.

[276] Wang, Gang-Jin, Xie, Chi, and Chen, Shou. "Multiscale correlation networks analysis of the US stock market: a wavelet analysis". In: *Journal of Economic Interaction and Coordination* 12.3 (2017), pp. 561–594.

[277] Wang, Gang-Jin, Xie, Chi, and Stanley, H Eugene. "Correlation structure and evolution of world stock markets: Evidence from Pearson and partial correlation-based networks". In: *Computational Economics* 51.3 (2018), pp. 607–635.

[278] Wang, Gang-Jin et al. "Random matrix theory analysis of cross-correlations in the US stock market: Evidence from Pearson's correlation coefficient and detrended cross-correlation coefficient". In: *Physica A: Statistical Mechanics and its Applications* 392.17 (2013), pp. 3715 –3730. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2013.04.027. URL: http://www.sciencedirect.com/science/article/pii/S0378437113003403.

[279] Wang, Gang-Jin et al. "Statistical Properties of the Foreign Exchange Network at Different Time Scales: Evidence from Detrended Cross-Correlation Coefficient and Minimum Spanning Tree". In: *Entropy* 15.12 (2013), 1643–1662. ISSN: 1099-4300. DOI: 10.3390/e15051643. URL: http://dx.doi.org/10.3390/e15051643.

[280] Warton, David I. "Penalized Normal Likelihood and Ridge Regularization of Correlation and Covariance Matrices". In: *Journal of the American Statistical Association* 103.481 (2008), pp. 340–349. DOI: 10.1198/016214508000000021. eprint: https://doi.org/10.1198/016214508000000021. URL: https://doi.org/10.1198/016214508000000021.

[281] White, Douglas R and Reitz, Karl P. "Graph and semigroup homomorphisms on networks of relations". In: *Social Networks* 5.2 (1983), pp. 193–234.

[282] Wilcox, Diane and Gebbie, Tim. "On the analysis of cross-correlations in South African market data". In: *Physica A: Statistical Mechanics and its Applications* 344.1 (2004). Applications of Physics in Financial Analysis 4 (APFA4), pp. 294 –298. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2004.06.138. URL: http://www.sciencedirect.com/science/article/pii/S0378437104009537.

[283] Wiliński, M. et al. "Structural and topological phase transitions on the German Stock Exchange". In: *Physica A: Statistical Mechanics and its Applications* 392.23 (2013), pp. 5963 –5973. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2013.07.064. URL: http://www.sciencedirect.com/science/article/pii/S037843711300695X.

[284] Xia, Lisi et al. "Comparison between global financial crisis and local stock disaster on top of Chinese stock network". In: *Physica A: Statistical Mechanics and its Applications* 490 (2018), pp. 222–230.

[285] Yan, C. et al. "Power–law properties of Chinese stock market". In: *Physica A: Statistical Mechanics and its Applications* 353 (2005), pp. 425 –432. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2005.02.010. URL: http://www.sciencedirect.com/science/article/pii/S0378437105001469.

[286] Yan, Xin-Guo, Xie, Chi, and Wang, Gang-Jin. "Stock market network's topological stability: Evidence from planar maximally filtered graph and minimal spanning tree". In: *International Journal of Modern Physics B* 29.22 (2015), p. 1550161. DOI: 10.1142/S0217979215501611. eprint: https://doi.org/10.1142/S0217979215501611. URL: https://doi.org/10.1142/S0217979215501611.

[287] Yan, Yan et al. "Development of Stock Networks Using Part Mutual Information and Australian Stock Market Data". In: *Entropy* 22.7 (2020), p. 773. ISSN: 1099-4300. DOI: 10.3390/e22070773. URL: http://dx.doi.org/10.3390/e22070773.

[288] You, Tao, Fiedor, Paweł, and Hołda, Artur. "Network Analysis of the Shanghai Stock Exchange Based on Partial Mutual Information". In: *Journal of Risk and Financial Management* 8.2 (2015), 266–284. ISSN: 1911-8074. DOI: 10.3390/jrfm8020266. URL: http://dx.doi.org/10.3390/jrfm8020266.

[289] Zalesky, Andrew, Fornito, Alex, and Bullmore, Ed. "On the use of correlation as a measure of network connectivity". In: *NeuroImage* 60.4 (2012), pp. 2096 –2106. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2012.02.001. URL: http://www.sciencedirect.com/science/article/pii/S1053811912001784.

[290] Zhang, Yiting et al. "Will the US economy recover in 2010? A minimal spanning tree study". In: *Physica A: Statistical Mechanics and its Applications* 390.11 (2011), pp. 2020 –2050.

[291] Zhao, Longfeng, Li, Wei, and Cai, Xu. "Structure and dynamics of stock market in times of crisis". In: *Physics Letters A* 380.5 (2016), pp. 654 –666. ISSN: 0375-9601. DOI: https://doi.org/10.1016/j.physleta.2015.11.015. URL: http://www.sciencedirect.com/science/article/pii/S0375960115009937.

[292] Zhou, Hongding and Slater, Gary W. "A metric to search for relevant words". In: *Physica A: Statistical Mechanics and its Applications* 329.1 (2003), pp. 309 –327. ISSN: 0378-4371. DOI: https://doi.org/10.1016/S0378-4371(03)00625-3. URL: http://www.sciencedirect.com/science/article/pii/S0378437103006253.

[293] Zhuang, R., Hu, B., and Ye, Z. "Minimal spanning tree for Shanghai-Shenzhen 300 stock index". In: *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. 2008, pp. 1417–1424.

[294] Zou, Hui and Hastie, Trevor. "Regularization and variable selection via the Elastic Net". In: *Journal of the Royal Statistical Society, Series B* 67 (2005), pp. 301–320.

[295] Zou, Hui, Hastie, Trevor, and Tibshirani, Robert. "On the "degrees of freedom" of the lasso". In: *Ann. Statist.* 35.5 (Oct. 2007), pp. 2173–2192. DOI: 10.1214/009053607000000127. URL: https://doi.org/10.1214/009053607000000127.