Network Modeling of Extreme Dependence in High-Dimensional Financial Time Series

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Siddharth G J Mohapatra

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

May, 2023

Supervisor: Rituparna Sen

Certificate

This is to certify that this dissertation entitled Network Modeling of Extreme Dependence in High Dimensional Financial Time Series towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Siddharth G J Mohapatra at Indian Institute of Science Education and Research under the supervision of Rituparna Sen, Associate Professor, Indian Statistical Institute Bangalore, during the academic year 2022-2023.

Rituparna Sen

Committee:

Rituparna Sen

Anindya Goswami

This thesis is dedicated to IISER Pune.

Declaration

I hereby declare that the matter embodied in the report entitled Network Modeling of Extreme Dependence in High-Dimensional Financial Time Series are the results of the work carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Rituparna Sen, ISI Bangalore, and the same has not been submitted elsewhere for any other degree.

Siddharth Gita
Jayanta Mohapatra

Siddharth G J Mohapatra

Acknowledgments

This thesis would not have been possible without the guidance and support of my supervisor, Dr Rituparna Sen. She has taught me so much about the science and art of statistics. Her guidance and advice were truly invaluable. I could not have asked for a better mentor. I would also like to thank Dr Sumanta Basu for his help and counsel, especially with the programming part of the project.

The past year was incredibly enjoyable and transformative for me, both as a person and as a scientist. And for this, I have two special people to thank: Suryadeepto, who inspired me with his passion for science and for this world, and Shivani, who inspired me with her love, grit and determination.

This section could not be complete without expressing my gratitude to my parents, who provided valuable funding in addition to their constant love and unconditional support.

I would like to thank the fraternity of Hostel-1 and the entire community of IISER Pune for making my final year so memorable. Perhaps the real research is the friends we make along the way.

Lastly, I want to thank you, the reader, for your interest in my work. It means the world to me.

Abstract

Modelling the systemic risk of the ever-growing global financial market has garnered significant interest since multiple worldwide crises showed that contagion of financial distress could occur on a system-wide scale. In this thesis, We study the phenomenon of extreme dependence, also known as financial contagion, in which extreme events in financial time series data move together and propagate distress. To model the interconnected relationships of a large market, where the dimension is high, we employ network-based methods. In particular, we construct an original framework based on the theory of generalized linear models to detect contagion by quantifying the amount of co-movement of extreme events. We employ the LASSO to perform a simultaneous estimation of the links of a specific instrument to the rest of the market. Finally, we use the estimates to represent the market as a network and derive summary statistics that illustrate relevant facets of the underlying complex network, such as the mean connectivity of the network. We test the validity of this method by comparing the time series of these summary statistics against known crisis periods.

Contents

Chapter 1

Introduction

Network modelling has been widely employed in financial mathematics in recent years to study the properties arising from the interconnectedness of markets, as network based methods provide a straightforward way to describe the numerous pairwise relationships occurring within a financial market [1]. We are particularly interested in the network modelling of extreme dependence of financial time series, as the phenomenon of extreme dependence is a proxy for financial contagion. This is in contrast to the correlation of time series which may be inflated due to co-movement during tranquil periods. We seek to extend the method of Residual and Recurrence Times developed in [2] in order to address the system-wide nature of contagion and provide a direct path to estimate and interpret connection strength between financial entities by employing generalized linear models, particularly when the dimension is high. We analyse various network summary statistics and see how they correlate with known crisis periods. This gives insight into how the network representation of markets changes during periods of financial distress and hence forms an alternate characterization of contagion, which could enhance the assessment of systemic risk and prediction of crises.

1.1 Motivation

The rise of globalization and the advent of the Internet have made financial markets around the world more interconnected than ever before. Prices of financial entities respond to information gathered from all over the globe and change in the scale of nanoseconds. However,

catastrophic events such as the 2008 US Subprime Crisis have demonstrated that this interconnectedness can lead to transmission of financial shock across continents and induce crisis periods that reflect in the time series data of a multitude of financial entities. This exhibition of interdependence is termed 'financial contagion'.

Consequently, much interest has been shown for understanding and measuring the systemic risk associated with financial contagion. It became important to identify highly interconnected entities, since these entities pose a higher risk to the entire system owing to the multiple possible channels of contagion [3]. A natural structure of interest in such a system where many co-moving quantities are interconnected is a network of these said quantities. Networks provide an intuitive way of visualizing the many relationships between these quantities, as well as providing a clear and direct way of quantifying the overall connectivity of the system of financial institutions, and identifying channels of contagion. We can subsequently derive summary network measures as proxies for systemic risk.

One can imagine how the information of a 'crisis' spreads throughout markets in a simple way: market participants react to 'extreme events' in the time series of affected financial entities by making trades that cause further 'extreme events' to occur in related entities. A simple and somewhat naive method to assess the connections between institutions and entities is to consider the statistical correlation of the time series of these entities. Indeed, studies have found increased correlations between developed markets after crashes [4] [5] [6]. However, simple correlation values could be high even during non-crisis periods where extreme events do not move together.

Therefore, we would like to consider 'extreme dependence' and study the interdependence and transmission of these extreme events taken separately. Extreme dependence is of interest because it is a more unambiguous indication of the contagion of crises. as we see abnormal events occurring and then being transmitted, as expected in a crisis [2].

1.2 Survey of Literature

The study of the application of graph theoretic techniques to analyse financial data was pioneered by Mantegna in 1999 [7], in a paper where he demonstrated that the Pearson's

correlation between stock returns in the Dow-Jones Industrial Average captured information about the 'hierarchy' of stocks. The research in the recent decades has been mainly focused on this theme of extracting relevant economic information from network representations of a financial markets. These network-based methods have been extended to entities like stocks, currencies and debt instruments [8] [9].

The broad themes of the recent theoretical literature of network modelling of financial data pertain to the construction of statistically robust hypothesis testing and using system-wide measures for estimation of connections, along with quantitative characterization of the properties of these networks. In [3], the authors establish a sound, system-wide framework for the statistical testing of hypothesis of relationships between entities. This approach is based on VAR modelling and Granger causality [10], which are widely used in literature [11] [12] [13] [14] [15]. This is combined with the use of generalized linear models to obtain quantititative estimates of links in the network. Unlike correlation based networks [8] [16] [17] found in earlier literature, causality based graphs can be directed and hence are more appropriate for detecting relationships in financial data. To address systemic risk, new theoretical formulations such as conditional Value-at-Risk [18], contagion mapping [19] and systemic expected shortfall [20] have been developed. To explicitly encourage sparsity in the generated complex networks, a variety of subset selection strategies are employed in literature, however their stability in the high dimensional setting is questionable [21]. Shrinkage methods such as ridge regression and LASSO are found to be more robust, and provide consistent estimates [22] [23]. Theoretical formalisation of the space of network representations that allow us to compare to assign a geometric nature to this space have also been developed [24].

On the empirical frontiers in literature, we find an ever-increasing methods for estimating the linkages between financial series. In addition to stock price data [2] [3], volatility [14] [25] and bank data [12] [26] [27] are commonly used to build networks and analyse co-dependencies of entities. Generally, authors prefer to use weekly data instead of daily closing prices to insulate against spillover effects [28] [2], and then segregate the data according to geography [13] [29]. Granger causality along with autoregressive vector modelling is again quite popular, due to its simplicity and ease of interpretation [30] [13] [14]. Other approaches to achieve the same use partial correlation [31], copula estimation [28] and statistical co-integration [30]. The data used in literature is interspersed with known crisis periods, most prominently featuring the 2008 global financial crisis [12] [14] [32] [33].

The generated complex networks found in literature form a considerable variety. For instance, both directed [13] [14] and undirected [30] [28] networks are abundant. The weights of the links are omitted when the primary purpose of network modelling is to uncover clusters and identify highly connected nodes [30] [13]. To filter out the most important financial relationships in a generated network, the most common strategies used are to extract the minimal spanning tree (MST) [7], or some variation thereof. Recently, planar maximally filtered graphs (PMFG) [13] have been employed for this purpose. The set of network characteristics under analysis is diverse: average degree [3], clustering [28], sparsity [14], and betweenness centrality [30] are some standard network characteristics analysed in literature due to ease of interpretation and because they help in identification of nodes most pertinent to the transmission of financial stock.

1.3 Original Contribution

In this thesis, we review the method of Recurrence and Residual Times (RRT) [2] for testing extreme dependence in financial data and assess the model, based on trends in current literature. While the RRT is able to establish evidence of directed contagion in bivariate time series, i.e., pairwise relationships, we find that it has notable limitations which prevent accurate system-wide estimation of the relationships in financial data.

Our contribution to literature is detailed as follows: based on these limitations, we modify the method and introduce a new system-wide, regression-based method for network modelling. We develop the theoretical framework and use it to analyse the network characteristics of two datasets of global index data. Both of these datasets contain the infamous 2008-09 financial crisis, a period of intense global contagion across the globe, and hence serve as a good test of the new method.

Chapter 2

Preliminaries

All definitions follow standard conventions in modern literature.

Definition 2.0.1. ***Network***
A directed, weighted network or graph is a set $G = (V, E, W)$ *where* V *is the set of vertices or nodes,* $E = \{e_{ij} = (v_i, v_j) : v_i, v_j \in V, i \neq j\}$ *is the set of ordered pairs of edges or links, and* W *is the set of weights such that* w_{ij} *is the weight of the directed edge* (v_i, v_j)*.*

Definition 2.0.2. ***Adjacency Matrix of Network***
The adjacency matrix A *for a weighted, directed network with* n *nodes and with no loops (i.e. no edges from a node to itself), is a square* $n \times n$ *matrix such that the* (i, j)*-th entry*

$$A_{ij} = \begin{cases} w_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where $w_{ij} \in \mathbb{R}$ *is the weight of the directed edge from node* i *and node* j*. While it is symmetric for undirected networks, it may not be for directed ones.*

Definition 2.0.3. ***Degree of a node***
The average in-degree $d_{in,i}$ *of a node* i *in a directed graph* G *is defined as*

$$d_{in,i} = card(\{e_{ji} : j \in V, j \neq i\})$$

Similarly, the out-degree $d_{out,i}$ of a node i in a directed graph G is defined as

$$d_{out,i} = card(\{e_{ij} : j \in V, j \neq i\})$$

Definition 2.0.4. ***Average degree of a network***
In case of an undirected network, the average degree is simply given by degree sum divided by the total number of nodes.

$$\bar{d} = \frac{\sum_{i \in V} d_i}{n} = \frac{2e}{n}$$

where n is the total number of nodes and e is the total number of edges.
The average in-degree $\bar{d_{in}}$ of a directed graph G is defined as

$$\bar{d_{in}} = \frac{\sum_{i \in V} d_{in,i}}{n} = \frac{e}{n}$$

where n is the total number of nodes. The average out-degree can be defined similarly, and is equal to the average in-degree, as each edge contributes equally to both the in-degree sum and the out-degree sum. The average degree for an directed graph is thus

$$\bar{d_{direct}} = \frac{2e}{n}$$

The average degree of a network represents the level of connectivity in it, and describes how many channels are available, on average, for information to be propagated in the network.

Definition 2.0.5. ***Clustering coefficient***
The local clustering coefficient for directed graphs is defined as

$$C_i = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i\,(k_i - 1)}$$

where k_i is the number of vertices, $|N_i|$, in the neighbourhood, N_i, of a vertex. The neighbourhood N_i for a vertex v_i is defined as the set of its immediately connected nodes:

$$N_i = \{v_j : e_{ij} \in E \vee e_{ji} \in E\}\,.$$

The network average clustering coefficient is defined as:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

The clustering coefficient is a measure of how crowded or dense the graph is.

Definition 2.0.6. ***Transitivity***
Transitivity, also known as global clustering coefficient, is defined as the ratio of the triangles and the connected triples in the network. It is computed by taking the local clustering coefficient for each node i and weighing it by a factor $k_i(k_i - 1)$, where k_i is the number of neighbours of node i. Compared to the network average clustering coefficient, it places a higher weight on nodes with high degree.

Definition 2.0.7. ***Average Shortest Path***
A directed path between two nodes in a digraph is a sequence of edges leading from one node to the other in a certain direction (from source to sink). The weight of a path is defined as the sum of the weights of its edges.

The average shortest path of a network is defined as the average of the length of all shortest paths in the networks. This can be weighted as well, in which case it is called weighted average shortest path. The average shortest path can be thought of as a measure of how quickly information can travel through the network.

Definition 2.0.8. ***Betweenness Centrality***
The normalized betweenness centrality of node i can be defined as BC_i

$$\mathrm{BC}_i = \frac{\sum_j^n \sum_k^n g_{jk}(i)/g_{jk}}{N^2 - 3N + 2}, \quad j \neq k \neq i, \quad j < k$$

where $g_{jk}(i)$ is the number of shortest paths between vertex j and vertex k that pass vertex i, and g_{jk} is the total number of shortest paths between vertex j and vertex k.
Betweenness centrality assigns a numerical importance to the transmission ability of each node and allows us to filter out the most important nodes for transmission of shock in the network.

Definition 2.0.9. ***Extreme Event***
For a given time series X_t, an extreme event is defined to be an event above or below a

chosen threshold. A common way of choosing this threshold is by using percentile of the empirical distribution of X_t. For example, an upper threshold might be the $95th$ percentile, and similarly, a lower threshold may be chosen to be the $5th$ percentile.

Definition 2.0.10. ***Recurrence Time***

The recurrence time V_j is defined simply as the time lag from one extreme event to the next one in a given time series Y.

Definition 2.0.11. ***Raw Residual Time***

The time lag between an extreme event in series Y and the next extreme event in series X is defined to be the Raw Residual time and denoted as Z_k. We can formalize this definition by writing:

$$Z_k = \sum_{i=1}^{N} U_i - \sum_{j=1}^{k} V_j + 1,$$

where U_i are the recurrence times for X, V_j are the recurrence times for Y, and $N = \underset{n}{arg\,min} \left\{ \sum_{i=1}^{n} U_i - \sum_{j=1}^{k} V_j \geq 0 \right\}$.

The "+1" on the right hand side of the definition means that if two extreme events occur at the same time, we consider the transmission time to be 1 by convention.

Definition 2.0.12. ***Residual Time***

For target series X given another series Y, the sequence of residual times, denoted by $\{W_k\}$, is constructed by taking the sequence of Raw Residual times $\{Z_k\}$ and then eliminating the overlaps. Mathematically, $\{W_k\} = \{Z_k\} \setminus \{Z_{k'} \ : \ Z_{k'} + V_{k'} = Z_{k'-1}\}$

Chapter 3

Methods

In this chapter, we review the RRT method method for testing financial contagion between financial time series. We then build upon the method to create an original method for modelling the underlying network of inter-dependencies in financial data.

We are interested in modelling the relationships between the various pairs of financial entities in a given market as a complex network, in which these entities are represented as nodes. Specifically, we want to establish a weighted edge between two nodes if the extreme events in the associated time series data (e.g. log-returns series, volatility series, etc.) have co-movement that cannot be explained by pure chance. The methods discussed in this chapter represent two attempts at obtaining a complex network that represents the state of a financial market based on the movements and interactions of extreme events occurring within it.

3.1 The RRT Method

The RRT method provides a novel way of detecting periods of time when extreme events move together in bivariate time series by comparing the distributions of recurrence and residual times between the two given series [2]. Specifically, the hypothesis for directed contagion between these two time series is statistically tested by the RRT Method. We shall first discuss the RRT Method in detail before motivating a modified version of it, wherein we

use the idea of residual times of a series with respect to another influencing the recurrence times of that series to formulate a regression based framework for network construction.

3.1.1 Details of the Method

We begin by noting definitions of residual and recurrence times (see Definitions 2.0.10, 2.0.11, 2.0.12). The basic idea is as follows: we consider the bivariate time series consisting of two components, time series X_t and Y_t, both of which belong to a larger set of time series $\{\mathcal{S}_i\}$. We filter the extreme events associated with these two series using a chosen percentile value, which leads us to two new sequences of extreme events. The residual and recurrence times are then computed for the sequences of extreme events. If there is no contagion to series X_t from the series Y_t, then the distribution of recurrence times of extreme events of X_t will be the same as the residual times of the extreme events of X_t with respect to Y_t. Intuitively, this can be interpreted as asserting that the extreme events in Y_t have no influence on the occurrence of the extreme events in X_t, and hence the respective distributions are uncoupled. Similarly, contagion in the other direction, i.e., from series X_t to the series Y_t can also be tested.

The algorithm implementing this idea uses a permutation test to compare the distributions of recurrence and residual times. Under the null hypothesis of no contagion, we would expect these distributions to be the same. The raw residual times and recurrence times are clubbed into one single set from which two subsets are randomly drawn. The corresponding residual time is computed, which then allows us to calculate a test statistic. When the previous steps are iterated over, we obtain the distribution of the test statistic. Finally, the p-value of this test statistic is computed and can be used to straightforwardly quantify the amount of directed contagion.

3.1.2 Network Generation using the RRT Method

To generate a weighted network representation of a financial market using the RRT Method, we must first decide upon the process of assigning a weight to the edge drawn between two nodes. In his seminal paper, Mantegna used the Pearson correlation ρ_{ij} to define a "distance" d_{ij} between two nodes v_i and v_j, and showed that this choice of "distance" defines a metric

on the network [7]. The distance between two nodes effectively acts as the weight to the edge drawn between these two nodes. The exact formula used is:

$$d_{ij} = 2\sqrt{1 - \rho_{ij}} \tag{3.1}$$

This distance function is an illuminating choice of edge weight. We first note that for any two time series, $\rho_{ij} \in [-1, 1]$, which implies that $d_{ij} \in [0, 2\sqrt{2}]$, guaranteeing that the distance between any two nodes $d_{ij} \geq 0$, with 0 being the case where the two time series are completely correlated, i.e. they are the same time series up to scale and location transformations. Secondly, it can be easily seen that the distance function is a decreasing function of the correlation coefficient in the interval $[-1, 1]$. This essential property ensures that nodes with higher correlation are placed near each other in the network and thus, a set of highly correlated nodes will form a close cluster in the network.

The natural and naive choice of edge weight would be to simply consider the p-value of the test statistic from the RRT Method, since a low p-value (for example, if $p < 0.001$) corresponds to a high degree of extreme dependence and would lead to clustering in the network by placing strong dependent entities close. The complex network thus generated represents the state of the market in terms of contagion.

Therefore, the formula for constructing the required directed adjacency matrix is:

$$\mathcal{A}_{ij} = \begin{cases} w_{ij} = 2\sqrt{1 - P_{ij}} & \text{, if } i \neq j \\ 0 & \text{, if } i = j \end{cases} \tag{3.2}$$

where P_{ij} is the p-value of the test statistic measuring directed contagion from the node j to the node i. The major drawback of this choice of edge weight is that it would lead to the creation of a dense graph, with all possible edges being drawn. A simple method to introduce sparsity in the network would be to consider a statistical significance level α, which can be used as a cut-off value to filter out weak and unimportant relationships in the data. The network thus constructed can be further refined by filtering out insignificant relationships among the nodes by extracting an underlying graph structure, like the minimal spanning tree (MST) or the planar maximally filtered graph (PMFG).

When this network generation algorithm is repeated over a sliding window that moves along

the dataset, it produces a time varying complex network. Consider one such network $\mathcal{N}_{\sqcup}$, with the set of nodes $\{v_i\}$ and the edge set $\{e_i\}$. We can now compute the time series of *network characteristics*, like average degree or the clustering coefficient (Definitions 2.0.4, 2.0.5). These new time series provide insight on the state of the network over a period and help in the identification of system-wide crisis periods. For example, the average degree of the network, which intuitively represents the average connectivity in a network, would be expected to rise during crisis periods when contagion is widespread.

Additionally, important nodes that contribute most to contagion can be identified trivially: the most connected nodes. Furthermore, using the betweenness centrality (Definition 2.0.8) can help identify nodes that participate in pathways for contagion and serve as a propagator of financial distress.

3.2 The Regression based RRT Method

The RRT method locates time periods where extreme events move together and is thus able to identify crisis periods based on the distribution of the recurrence and residual times of extreme events that are dispersed throughout a set of given time series. However, this approach is has a few limitations which we seek to address. The modifications suggested by the existence of these limitations leads up towards a regression based method for analyzing the time series data of financial markets. We then explain the mathematical model underlying the method, and finally, its implementation.

3.2.1 Motivation for a Modified RRT Method

We first begin by noting the most prominent drawbacks of the RRT Method:

1. **Pairwise estimation vs System-wide estimation**: The accurate estimation of edge weights in complex networks is crucial for various applications such as predicting contagion, identifying influential nodes, and understanding network dynamics. The

RRT method compares distributions to assign a weight to the edge drawn between two nodes. This has two drawbacks: first, it demands n^2 estimations corresponding to all the n^2 possible edges in a network with n nodes, which incurs significant computational costs. Secondly, it is vulnerable to spurious correlations which might creep into the estimates, which will produce a detrimental effect on their accuracy. To illustrate, consider a set of three time series of stock prices. Let us call them A, B and C. If the true network is such that the value of C drives both A and B, i.e. C is strongly linked to both A and B, then significant correlation will be found in the latter two if pairwise estimation is performed. However, the truth is that they are weakly linked and show spurious correlation due to being driven by the same underlying stock, C. Therefore, we are prompted to employ a system-wide estimation strategy which will avoid these problems by simultaneous estimation of all the edge weights in a complex network in a given window of time. This amounts to going from the bivariate estimation scheme of the RRT Method to the multivariate regime. This is achieved by modelling the phenomenon of contagion using generalized linear models to estimate the edge weights.

2. **Sparsity and the small world network**: The assumption of sparsity in a complex network that depicts real world relationships is based on empirical observations [34]. The structure of complex networks that depict real-world relationships often follows the small-world network paradigm, with a scale-free distribution of degrees given by a power law of the form $P_{deg}(x) = c \cdot x^{-\alpha}$, where α is a scaling parameter and c is the constant of normalization [35] [36]. This might be because networks in the real world grow by preferential linkage. To account for this assumption of sparsity, we need to select a subset of all possible edges which would convey the most important relationships in the network, meaning that the number of edges in the final estimated network, denoted by e, should be much smaller than n^2.

3. **Interpretation of edge weights**: In the case of the RRT Method, edges are assigned weights equal to a p-value made from comparing distributions, which is harder to interpret. On the other hand, regression coefficients as edge weights have a very straightforward interpretation: they represent the strength and direction of influence that one entity will have on the other, in some scale. This can be quantified by analysing the specific model used for estimation. For example, logistic regression coefficients have the interpretation of representing the change in log-odds of success associated with a

unit increase for each predictor.

These limitations inspire us to move towards a system-wide regression based approach for modelling relationships in financial data. To accommodate for the assumption of sparsity and to move towards a system-wide scheme of estimation, we compute all the coefficients associated with each node simultaneously and also select a subset of predictors. This is achieved by employing the LASSO (Least Absolute Shrinkage and Selection Operator) [37].

3.2.2 Modelling Approach

The Setup

We start by describing the setup of the model. We are given a set of n time series $\{\mathcal{S}_i\}$, $i = 1, 2, \ldots, n$, all of which describe the movement of a certain quantity of interest (return of assets, trading volume, volatility, etc.) for financial entities (stocks, derivatives, etc.) in a financial market. The task is to construct a network model of the market from these time series based on the relative quantities of extreme dependence between the n entities.

The most natural quantity to consider is the spot price of a financial instrument, but this is not always the case. For example, in the case where the financial entity is a stock in a share market, whose price at time t is denoted by P_t, then a basic quantity is the log-return $r(t)$ of the asset, defined as:

$$r(t) = \log \frac{P_t}{P_{t-1}} \tag{3.3}$$

The log-return series is widely studied in mathematical finance instead of the spot price series, as it presents certain advantages. Firstly, the log-returns are additive, hence the log-returns for time period $t = t_i$ to $t = t_f$ can be computed simply:

$$r(t_f) = \sum_{t=t_i}^{t=t_f} r(t) = \log \frac{P_{t_f}}{P_{t_i}} \tag{3.4}$$

Secondly, a common assumption for prices is that they are distributed log-normally, which leads to a normal distribution for the log-returns. Motivated by these reasons, we consider the log-returns of prices in our statistical analysis of data, in which case the set of time series

$\{\mathcal{S}_i\}$ represents the log-returns of n assets in a market over time. However, we emphasize that this setup is general and that the set $\{\mathcal{S}_i\}$ could represent quantities like volatility or trading volume.

Since we are interested in extreme dependence, we filter out the extreme events (Definition 2.0.9) and also note their time of occurrence for each series $\mathcal{S}_i$. The pairs $(\mathcal{S}_j(t_{exceed}), t_{exceed})$ of extreme values and their time of occurrence enable the computation of recurrence and residual times for a chosen index j with respect to any other index j'.

Let $\mathcal{Y}$ be a chosen series from the set of n time series of equal length $\{\mathcal{S}_i\}$, $i = 1, 2, \ldots, n$. Let the recurrence times of the extreme events of $\mathcal{Y}$ be given by the sequence $\{Y_k\}$. The length of this sequence depends on the chosen percentile threshold; lowering the threshold lengthens the sequence. For each extreme event in $\mathcal{Y}$, we have one corresponding recurrence time Y_i (Definition 2.0.10).

Let us denote the time length of the series $\mathcal{Y}$ by T, and the chosen threshold for extreme events as $\rho \in (0, 1)$. To illustrate, consider an example where $\rho = 0.05$. This corresponds to a percentile value of 5%ile, and implies that only the bottom 5% of points by magnitude are considered to be extreme events. On the other hand, if we wish to analyse the movements of extreme events in the other other direction, i.e., the top 5%ile of points, we choose $\rho = 0.95$. In general, when $\rho > 0.5$, then it represents an upper threshold and vice versa when $\rho < 0.5$. The number of extreme events $\ell \in \mathbb{N}$ in each series in the set $\{\mathcal{S}_i\}$, $i = 1, 2, \ldots, n$ is thus

$$\ell = \min\{\rho \times T, (1 - \rho) \times T\} \tag{3.5}$$

This formula also provides the length of the sequence $\{Y_k\}$ as each extreme event has exactly one corresponding recurrence time.

Model Assumptions

To apply linear regression to the data, we first need to make an assumption about the nature of the relationship of recurrence and residual times for a given series. Let us begin by

considering a simple case first. We start with a single time series, say X_t which takes values in $\mathbb{R}$. To define extreme events, we choose a *fixed lower* threshold given $\rho \in (0, 1)$, using which we can compute the recurrence times for X_t. Let $\{U_i\}$ be the sequence of recurrence times thus obtained. We can easily see that this sequence follows a Geometric distribution with the parameter $(1 - \rho)$ by computing the probability mass function of $U_i \in \mathbb{N}$:

$$\mathbb{P}(U_i = u) = (1 - \rho)^{u-1} \cdot \rho \tag{3.6}$$

This follows from the fact that an extreme event will be observed at a rate of ρ in case of upper thresholds. This allows us to interpret (1-ρ) as a probability of success in the classical setup for a Geometric distribution, where we seek the distribution of the number of failures before the first success in a series of Bernoulli trials. In the case of a fixed upper threshold, the analysis is the same with the success probability replaced by ρ instead of $(1 - \rho)$.

We have now established that the recurrence times will follow a Geometric distribution:

$$U_i \sim Geom(1 - \rho) \quad \forall i$$

This distribution has is supported on the the set of possible values u that the recurrence time U_i can take, and clearly this set is $\mathbb{N}$, the set of natural numbers.

Returning back to the original setup, we have the time series $\mathcal{Y}$ and its recurrence times $\{Y_i\}$. Consider an extreme event that occurs at time t_0 in $\mathcal{Y}$, and let its recurrence time be Y_{t_0}. Our goal is to quantify the strength of co-movement of the extreme events in $\mathcal{Y}$ with respect to the extreme events in the remaining $n - 1$ time series in the set $\{\mathcal{S}_i\} - \mathcal{Y}$. To do so, we introduce a modification in the definition of residual time from Definition 2.0.12.

Definition 3.2.1. ***Backward Looking Residual Time***
The sequence of residual times, denoted by $\{W_k\}$, for series X given another series Y is the time lag between an extreme event in X and the closest extreme event that occurred previously in Y.

In simpler terms, the Backward Looking Residual Times are effectively recurrence times of an extreme event in $\mathcal{Y}$, except that they do not measure the time lag between extreme events within $\mathcal{Y}$ itself, but from the closest past extreme events in the other $n - 1$ series.

We now have a complete setup for the application of linear regression. Our key assumption is that the recurrence times Y_i of $\mathcal{Y}$ can be modelled as a linear combination of some parameters and of the residual times with respect to the other series.

The response is the set of recurrence times $\{Y_i\}$, and the predictors are the backward residual times: for each of the ℓ recurrence times Y_i, we have a vector $\mathbf{x_i} = (1, x_{i1}, x_{i2}, \ldots, x_{i,n-1})$ of residual times from the remaining series in the set $\{\mathcal{S}_i\}$, $i = 1, 2, \ldots, n$. This can be interpreted as determining which financial entities are the strongest indicators of the occurrence of extreme events in the series $\mathcal{Y}$, which in turn determines which entities have extreme events that move with those of $\mathcal{Y}$.

We are now ready to state our assumptions:

1. We assume that the recurrence times of the extreme events can be modelled as a random variable Y_i which takes in $y_i \in \mathbb{N}$, being drawn from Geometric distributions with a mean $\mu \in [0, \infty)$.

2. We assume that the residual times of an extreme event in series i from another time series can be modelled as a random variable X_{ij} which takes values $x_{ij} \in \mathbb{N}$.

3. We assume that a linear predictor of residual times $X_{ij}\ \forall\ j$ is statistically correlated to a function of the mean of this distribution.

4. We assume that the underlying network of connections within the financial market is sparse.

Simple Linear Model

The simple linear model can be summarized mathematically as:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{n-1} x_{i,n-1} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, \ell \tag{3.7}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_{n-1})$ is the vector of coefficients, each encoding the dependence of Y_i on the $n-1$ remaining time series, and ε_i is the associated normally distributed random error such that $\mathbb{E}(\varepsilon_i) = 0\ \forall\ i$.

We can stack these ℓ equations and convert them into compact matrix notation as follows:

$$\mathbf{Y}_{\ell\times 1} = \mathbf{X}_{\ell\times(n-1)}\boldsymbol{\beta}_{(\mathbf{n-1})\times\mathbf{1}} + \boldsymbol{\epsilon}_{\ell\times\mathbf{1}} \tag{3.8}$$

where, $\mathbf{Y}$ is the response, $\mathbf{X}$ is the *design matrix*, $\boldsymbol{\beta}$ is the vector of coefficients and finally, $\boldsymbol{\epsilon}$ is the ℓ-dimensional random error with zero mean and constant variance, i.e.

$$\mathbb{E}(\boldsymbol{\epsilon}) = \mathbf{0}, \quad \mathrm{Cov}(\boldsymbol{\epsilon}) = \sigma^2\mathbf{I}_\ell$$

Suppressing the subscripts, we have:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_\ell \end{bmatrix} \tag{3.9}$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_\ell^\top \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1,n-1} \\ 1 & x_{21} & \cdots & x_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{\ell 1} & \cdots & x_{\ell,n-1} \end{bmatrix} \tag{3.10}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{n-1} \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_\ell \end{bmatrix} \tag{3.11}$$

Finally, we can write

$$\mathbb{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta} \tag{3.12}$$

since the expectation of the errors is zero.

Limitations of the Simple Linear Model

However, this model of simple linear regression is inappropriate for estimation of the strength of relationships between these time series. Note that in linear regression, the *linear predictor* for the i-th response $\eta_i = \sum_{k=0}^{k=(n-1)} \beta_k x_{ik}$ and the response itself vary over the set of real numbers $\mathbb{R}$, whereas in this setup, $Y_i \in \mathbb{N}$, due to which $\mathbb{E}(Y_i|\mathbf{X}) \in (0, \infty) \neq \mathbb{R}$. Additionally, the distribution of the recurrence times Y_i is Geometric, which is not compatible with a linear-response model, which requires that the response have normal distributions with constant variance. The model implies that the predictors influence the mean of the distribution $\mu = \mathbb{E}(\mathbf{Y}|\mathbf{X})$ by means of shock transmission. By design, a linear-response model demands that a constant change in the inputs will cause a constant change in the response. But, for a Geometric response, we would expect a change in predictors to reflect non-linearly in the response (a unit change in the predictors might cause an exponential change in the response). This problem carries over to the assumptions regarding the random errors, which might not hold for a Geometric response.

Therefore, we are prompted to change the model of regression to account for the non-linear nature of the problem. To model a linear relationship between the response and the predictors, given the Geometric nature of the response, we look towards the framework of Generalized Linear Models (GLMs).

3.2.3 Generalized Linear Regression Model for RRT

The central assumption of our model is that, in case of extreme dependence, the expectation of the i-the recurrence time of $\mathcal{Y}$ in the sequence $\{Y_k\}$ is dependent linearly on the $n-1$ backward residual times of its associated extreme event with respect to the other $n-1$ time series. We now know that this assumption is not adequate since the distribution of the response is Geometric, due to which the response takes discrete values from the set of natural numbers $\mathbb{N}$, which implies that its mean will always be positive.

In Generalized Linear Models, the response is allowed to take values from a distribution that belongs to the exponential family. The mean of this distribution, $\boldsymbol{\mu}$, is linked to the predictors by a link function $g(\cdot)$:

$$\mathbb{E}(\mathbf{Y} \mid \mathbf{X}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta}) \tag{3.13}$$

By applying the function $g(\cdot)$ on both sides, we obtain

$$\implies g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\eta} \tag{3.14}$$

where $\boldsymbol{\eta}$ is the linear predictor.

The matrix $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$ is called the *linear predictor*. In this new setup, while a constant change in the predictors produces a constant change in the function $g(\boldsymbol{\mu})$, the effect on μ is not restricted to be linear. Moreover, we can map the range of the mean of the response to $\mathbb{R}$ with an appropriate choice for the link $g(\cdot)$, which now makes the model suitable for ordinary linear regression.

Calculating the Link Function

To derive the link function, we begin by considering the theory of GLMs with a fresh setup. We follow McCullagh's book on the subject [38] to analyse the GLM with Geometric response.

Once again we start with the response vector $\mathbf{Y} = (y_1, y_2, \dots y_n)$. Do note that n here is different from the rest of the chapter, where it is instead replaced by ℓ since n is reserved for the total number of time series in the financial system. Each component $y_i \in \mathbb{N}$ of $\mathbf{Y}$ will have an distribution from the exponential family, and hence the mean $\mu_i \in [1, \infty)$. We can write the probability density function of Y in the general form as follows:

$$f_Y(y; \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \tag{3.15}$$

where, θ is the natural parameter, ϕ is the dispersion parameter, and $a(\cdot), b(\cdot)$, and $c(\cdot)$ are some functions that take specific forms depending on the distribution.

Next, we compute the log-likelihood for this distribution. Since the logarithm is a strictly increasing function over the reals, maximizing likelihood is the same as maximizing log-likelihood. Additionally, the distributions from the exponential family are logarithmically concave, which makes maximum likelihood estimation convenient. The likelihood function is given by:

$$\mathcal{L}(\theta, \phi; y) = f_Y(y; \theta, \phi) \tag{3.16}$$

$$l(\theta, \phi; y) = \ln \mathcal{L}(\theta, \phi; y) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \tag{3.17}$$

Next, let us consider the score function, which is the defined as the gradient of the log-likelihood with respect to the parameter vector:

$$s(\theta) \equiv \frac{\partial \ln \mathcal{L}(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \tag{3.18}$$

The expectation of the score conditional to the parameter θ can be evaluated simply:

$$\mathbb{E}(s \mid \theta) = \int_{\mathcal{X}} f(x; \theta, \phi) \frac{\partial}{\partial \theta} \ln \mathcal{L}(\theta, \phi; x) \mathrm{d}x \tag{3.19}$$

$$= \int_{\mathcal{X}} f(x; \theta, \phi) \frac{1}{f(x; \theta, \phi)} \frac{\partial f(x; \theta, \phi)}{\partial \theta} \mathrm{d}x \tag{3.20}$$

$$= \int_{\mathcal{X}} \frac{\partial f(x; \theta, \phi)}{\partial \theta} \mathrm{d}x \tag{3.21}$$

where $\mathcal{X}$ denotes the sample space for the distribution. We can now use the Leibniz rule to interchange the integral and differential operations, if we assume certain regularity conditions. The expression simplifies to:

$$\mathbb{E}(s \mid \theta) = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f(x; \theta.\phi) \mathrm{d}x = \frac{\partial}{\partial \theta}(1) = 0 \tag{3.22}$$

This leads us to the convenient relation that

$$\mathbb{E}\left(\frac{\partial l}{\partial \theta} \right) = 0 \tag{3.23}$$

Let us apply this to Equation 3.17. We have,

$$\frac{\partial l}{\partial \theta} = \frac{y - b'(\theta)}{a(\phi)} \tag{3.24}$$

where $b'(\theta) \equiv \dfrac{\mathrm{d}}{\mathrm{d}\theta} b(\theta)$. Using Equation 3.23, and noting that $b'(\theta)$ and $a(\phi)$ do not transform under expectation, we can write:

$$0 = \mathbb{E}\left(\frac{\partial l}{\partial \theta}\right) = \mathbb{E}\left(\frac{y - b'(\theta)}{a(\phi)}\right) = \frac{\mu - b'(\theta)}{a(\phi)} \tag{3.25}$$

Finally, we obtain the key result pertaining to the relationship between the mean μ of the distribution of Y and the natural parameter θ:

$$\mathbb{E}(Y) = \mu = b'(\theta) \tag{3.26}$$

Equation 3.26 gives us a template to find the link between the natural parameter and the mean of the distribution from the form of the distribution for any distribution in the exponential family. We can now identify the canonical link $g(\cdot)$. Since the link function relates the natural parameter to the mean, we must have:

$$\theta = g(\mu) \tag{3.27}$$

By comparing Equation 3.26 with Equation 3.14, we finally obtain:

$$\theta = g(\mu) = g(b'(\theta)) \tag{3.28}$$

Applying $g^{-1}(\cdot)$ on both sides leads to

$$g^{-1}(\theta) = b'(\theta) \tag{3.29}$$

Therefore, the canonical link can be identified as $g^{-1} = b'$. This concludes the derivation. ■

Now, we return to our original setup. We have Y_i as the response; it is the recurrence time which follows a Geometric distribution. Let the parameter of this distribution be p_i. Hence, Y_i will have a probability mass function given by

$$\mathbb{P}(Y_i = y) = (1 - p_i)^{y-1} \cdot p_i \tag{3.30}$$

Taking the natural logarithm on both sides of the equation gives:

$$\ln(\mathbb{P}(Y_i = y)) = (y-1)\ln(1-p_i) + \ln(p_i) = y\ln(1-p_i) - (\ln(1-p_i) - \ln(p_i)) \tag{3.31}$$

Comparing this with Equation 3.17 results in the following identification:

$$\theta = \ln(1-p_i) \tag{3.32}$$
$$c(\cdot) = 0 \tag{3.33}$$
$$a(\phi) = 1 \tag{3.34}$$
$$b(\theta) = \ln(1-p_i) - \ln(p_i) \tag{3.35}$$

For the Geometric distribution, we can easily compute the mean $\mu_i = \mathbb{E}(Y_i)$ of the distribution:

$$\begin{aligned}
&\mathbb{E}(Y_i) = \sum_{k=1}^{\infty} k \cdot p_i(1-p_i)^{k-1} \\
\implies &\mathbb{E}(Y_i) = p_i \sum_{k=1}^{\infty} k \cdot (1-p_i)^{k-1} \\
\implies &\mathbb{E}(Y_i) = p_i \left(-\frac{\mathrm{d}}{\mathrm{d}p_i} \sum_{k=1}^{\infty} (1-p_i)^k \right)
\end{aligned}$$

Since the success probability $p_i \in (0,1)$, the geometric sum converges to $\dfrac{1-p_i}{p_i}$

$$\begin{aligned}
\implies &\mathbb{E}(Y_i) = p_i \left(-\frac{\mathrm{d}}{\mathrm{d}p_i} \frac{1-p_i}{p_i} \right) \\
\implies &\mathbb{E}(Y_i) = p_i \left(\frac{\mathrm{d}}{\mathrm{d}p_i} \left(1 - \frac{1}{p_i} \right) \right) = p_i \left(\frac{1}{p_i^2} \right)
\end{aligned}$$

This gives us

$$\mathbb{E}(Y_i) = \frac{1}{p_i} = \mu_i \tag{3.36}$$

Combining this result with the identification that $\theta = \ln(1-p_i)$, we obtain:

$$\theta = \ln\left(1 - \frac{1}{\mu_i}\right) \tag{3.37}$$

This leads to the final identification for the canonical link function $g(\cdot)$:

$$g(\mu) = \ln\left(1 - \frac{1}{\mu}\right) \tag{3.38}$$

Remark Notice that while $\mu_i \in [1, \infty)$, the link function is undefined for the point $\mu = 1$. This is because an observed response of 1 is telling the model that extreme events happen next to each other, and provides the estimate of success probability $\hat{p}_i = 1$, which implies that the rate of extreme events is infinite.

The Final Model

Armed with the link function, we can finally state the model for a generalized linear regression of the data in terms of three components:

1. We have the response $\mathbf{Y}$, with components Y_i which represent the ith recurrence time of the time series $\mathcal{Y}$. Furthermore, the mean of the distribution is $\mu_i = \mathbb{E}(Y_i)$.
2. We have the matrix of linear predictors $\boldsymbol{\eta}$, formed by multiplying the design matrix with the coefficient vector: $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$
3. The link $g(\cdot)$ such that $g(\boldsymbol{\mu}) = \boldsymbol{\eta}$. For each component of the mean $\boldsymbol{\mu}$, this evaluates to $g(\mu_i) = \ln\left(1 - \frac{1}{\mu_i}\right)$.

Thus, we have the final model:

$$g(\mathbf{Y}_{\ell\times 1}) = \mathbf{X}_{\ell\times(n-1)}\boldsymbol{\beta}_{\mathbf{(n-1)\times 1}} + \boldsymbol{\epsilon}_{\ell\times\mathbf{1}} \tag{3.39}$$

where, as before, $\mathbf{Y}$ is the response, $\mathbf{X}$ is the *design matrix*, $\boldsymbol{\beta}$ is the vector of coefficients and finally, $\boldsymbol{\epsilon}$ is the ℓ-dimensional random error with zero mean and constant variance

Interpretation of Regression Coefficients

We have built our model to detect the strength of contagion from a series j to a target series i. In case contagion truly exists between these two series, it would imply the co-movement of

extreme events. This will be reflected in smaller residual times being associated with smaller recurrence times by means of the estimated coefficient $\hat{\beta}_j$. Therefore, the expected effect that an increase in the predictor X_j (the residual time) will have on the response Y_i (the recurrence time) is that it would make it longer.

Let us begin with a simple analysis to illustrate. We begin by explicitly writing the predicted success probability of the underlying Bernoulli trials that generate the distribution of the recurrence time Y_i, $\hat{p}_i$, as a function of the linear predictor, $\hat{\eta}_i = \sum_{k=0}^{n-1} x_{ik}\hat{\beta}_k$, calculated from data:

$$\hat{p}_i = 1 - \exp\left(\hat{\eta}_i\right) \tag{3.40}$$

This follows from the fact that the observed response y_i is the estimate for the mean of the distribution μ_i, which gives the success probability of the through Equation 3.36. The mean and the linear predictor are connected to each other via the link function, given by Equation 3.37.

If we assume that $\hat{\eta}_i$ is small enough so that $\hat{\eta}_i^2 << \hat{\eta}_i$, we can further simplify by ignoring the higher order terms in the Taylor expansion of the exponential function:

$$\hat{p}_i = 1 - \exp\left(\hat{\eta}_i\right) \tag{3.41}$$

$$\implies \hat{p}_i = 1 - \left(1 + \hat{\eta}_i + \frac{\hat{\eta}_i^2}{2!} + \cdots\right) \tag{3.42}$$

$$\implies \hat{p}_i \approx -\hat{\eta}_i \tag{3.43}$$

Now, let us analyse what happens if there occurs a change Δx_{ij} in the value x_{ij} of the predictor X_{ij}. The change in the linear predictor, $\Delta\hat{\eta}_i$ will be

$$\hat{\eta}_i + \Delta\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_{i1} + \ldots + \hat{\beta}_j \cdot x_{ij} + \hat{\beta}_j \cdot \Delta x_{ij} + \ldots + \hat{\beta}_{n-1} x_{i,(n-1)} \tag{3.44}$$

$$\implies \Delta\hat{\eta}_i = \hat{\beta}_j \cdot \Delta x_{ij} \tag{3.45}$$

But from Equation 3.43 that the change in the estimated success probability $\Delta\hat{p}_i$ can also be written as

$$\Delta\hat{p}_i \approx -\Delta\hat{\eta}_i \tag{3.46}$$

So finally, we can conclude that a change of unit value in the predictor would cause a change in the estimated success probability $\Delta\hat{p}_i$ as given by

$$\Delta\hat{p}_i \approx -\hat{\beta}_j \cdot \Delta x_{ij} \tag{3.47}$$

This analysis establishes the interpretation of the estimated coefficients: for a series j and target series i, the estimated coefficient $\hat{\beta}_j$ is the decrease in the predicted success probability $\hat{p}_i$ that will be caused by a unit change in the predictor X_{ij}. Ultimately, it means that if the sign of $\hat{\beta}_j$ is positive, we associate a decrease in the predictor to a decrease in the expected response Y_i. Thus, a highly positive estimate of the coefficient corresponds to a stronger case of contagion between the extreme events of the series i and j, while a negative estimate would imply the opposite. This motivates us to interpret the estimated coefficients as a measure of the "signal strength" of contagion between i and j.

3.2.4 Details of Implementation

To implement this GLM for network construction, we process the given n time series to obtain the sequences of extreme events for each series, and then compute the recurrence and residual times for each extreme event.

We iterate through each element of the set of given time series $\{\mathcal{S}_N\}, N = 1, 2, \ldots, n$. Let $\mathcal{S}_k$ be the time series under consideration in the current iteration. Then, the ℓ recurrence times of $\mathcal{S}_k$ form the response vector. This response is transformed element-wise by applying the link $g(s) = \ln\left(1 - \frac{1}{s}\right)$, since the observed value of the recurrence time $\hat{S_m}$ is the estimate for the mean of the underlying distribution $\hat{\mu_m}$. As remarked earlier, if the observed response $y_i = 1$, then we replace it the a value slightly more than 1 to avoid getting $\ln 0$. Using Equation 3.39, we run a regression and estimate the coefficient vector for the j-th series, and call it $\boldsymbol{\beta}_k$. This vector has the dimensions $(n-1) \times 1$.

Estimation Strategy

To compute the vector of coefficients $\boldsymbol{\beta}_k$, we might use Ordinary Least Squares (OLS) estimation. The OLS performs estimation by minimizing an objective function $H(\boldsymbol{\beta})$:

$$H(\boldsymbol{\beta}) = \sum_{i=1}^{\ell} \left| y_i - \sum_{j=1}^{n-1} X_{ij}\beta_j \right|^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2. \tag{3.48}$$

The estimate $\hat{\boldsymbol{\beta}}$ for the coefficients is thus:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{arg\,min} H(\boldsymbol{\beta}), \tag{3.49}$$

However, the use of OLS presents a few problems. In the high dimensional setting, where $\ell < n-1$, OLS fails as it requires the design matrix $\mathbf{X}$ to be full rank. Secondly, it will most likely produce non-zero estimates for all coefficients, thereby linking a series i to all other $n-1$ series. This will lead to the linkage of node i to all other $n-1$ nodes in the network, producing a dense network with n^2 edges, which is undesirable since we were motivated by the sparse nature of networks that represent the real world.

These problems are mitigated by abandoning the OLS scheme in favour of the LASSO (Least Absolute Shrinkage and Selection Operator) [37]. The LASSO solves the following problem:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{n-1}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \right\} \text{ subject to } \|\beta\|_1 \leq t, \tag{3.50}$$

where t is a specified parameter that controls the amount of regularizaion, and $\|u\|_p = \left(\sum_{i=1}^{N} |u_i|^p\right)^{1/p}$ refers to the ℓ^p norm. For added clarity, we can rewrite 3.50 in the Lagrangian form by transferring the constraint to inside the bracket, as follows:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^{n-1}} \left\{ \frac{1}{\ell} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\beta\|_1 \right\} \tag{3.51}$$

where ℓ is the number of samples. The optimal value for λ is found by varying it and choosing the value which produces the estimate with the least Mean Squared Error (MSE).

The LASSO encourages sparsity in the network by virtue of the constraint it places on the ℓ^1 of the coeffiecient estimate vector. This forces the weaker relations to go to zero,

while extracting the strongest links in the network. Additionally, the LASSO is robust even when the dimension is high, i.e., in the case where $\ell < n-1$. LASSO also enables estimation of the model with a notably smaller number of time points than the conventional approach of OLS, given that the underlying network is approximately sparse (which, if you recall, is one of our assumptions).

Hypothesis Testing

To construct a network based on the relationship between extreme events between time series, the basic hypothesis that we have to test has the following scheme: (null hypothesis) $\mathcal{H}_0$: there is no contagion from series $\mathcal{S}_i$ to the series $\mathcal{S}_j$ vs. (alternate hypothesis) $\mathcal{H}_1$: there exists no contagion from series $\mathcal{S}_i$ to the series $\mathcal{S}_j$.

Based on the GLM model, we can construct the hypothesis testing procedure for contagion from series $\mathcal{S}_i$ to the series $\mathcal{S}_j$ by consider the estimated coefficient $\hat{\boldsymbol{\beta}}_{ij}$: (null hypothesis) $\mathcal{H}_0 : \hat{\boldsymbol{\beta}}_{ij} < 0$ vs. (alternate hypothesis) $\mathcal{H}_1 : \hat{\boldsymbol{\beta}}_{ij} > 0$.

Notice that instead of the null hypothesis being $\mathcal{H}_0 : \hat{\boldsymbol{\beta}}_{ij} = 0$, its is $\mathcal{H}_0 : \hat{\boldsymbol{\beta}}_{ij} < 0$. The reason for this is that our model will provide a positive estimate for the existence of contagion, as a positive coefficient implies that the occurrence of extreme events is correlated in the two series under consideration. A negative estimate for the coefficient, however, will not imply contagion. It has the interpretation that an occurrence of extreme events in one series leads to *stabilization* in the target series, i.e., the movement of extreme events is anti-correlated.

The p-values for the coefficient estimates are provided by the statistical estimation software used, which in our case would be the `glmnet` [39] package.

3.2.5 Network Generation using the Modified RRT Method

We are now interested in constructing the weighted, directed adjacency matrix $\mathcal{A}$ with dimensions $n \times n$ for the network of financial instruments. Let i and j be two nodes of this network. Recalling Definition 2.0.2, we see that the (i, j)-th element of the adjacency matrix

is of the form.

$$A_{ij} = \begin{cases} w_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

where w_{ij} is the weight assigned to the edge from j to i.

The existence of negative estimates for the coefficients $\hat{\boldsymbol{\beta}}_{ij}$ means that we cannot interpret the edge weights as distances, since distances cannot be negative. Using a function to transform the weights to positive values (for example, the exponential) on the weights is also ineffective, since it will map zero to one, which will make the generated network possess full connectivity, disobeying the key assumption sparsity. Hence, we proceed with the natural and naive choice of edge weight by simply assigning the value of the estimated coefficient to the edge between nodes i and j:

$$w_{ij} = \hat{\boldsymbol{\beta}}_{ij}$$

Thus, the adjacency matrix $\mathcal{A}$ is defined case-wise as:

$$\mathcal{A}_{ij} = \begin{cases} \hat{\boldsymbol{\beta}}_{ij} & \text{, if } i \neq j \\ 0 & \text{, if } i = j \end{cases} \tag{3.52}$$

We summarise the implementation in Algorithm 1 below:

Algorithm 1: Building a Network with the Modified RRT Method

Input : Data, windowLen, step, percentile
Output: adjacency matrix $\mathcal{A}$

tolNum = floor((length(Data)-windowLen)/step)
ℓ = percentile*windowLen
for i **in** 1 to tolNum
do
| Compute exceedance times for set of time series $\{X_k\}_t$ in a window
| Compute recurrence and residual times for each time series X_k using exceedances
| Transform the response y_i of each slice of 3D array with link function $g(\cdot)$
| Perform LASSO on transformed response and $n-1$ predictors of length ℓ
| Use coefficients $\hat{\boldsymbol{\beta}}_{ij}$ as edge weights in matrix $\mathcal{A}$ and set diagonal to zero
end
return adjacency matrix $\mathcal{A}$ representing a weighted, directed network

Chapter 4

Analysis

In this section, we use the RRT and the modified RRT methods developed in the previous chapter to analyse two sets of financial data from global market systems. We process the raw time series of index data to compute log-returns. Using these set of log-returns series, we produce quantitative estimates of the inter-dependencies in financial markets. This information is then employed to draw a network representation of the market, generated by a taking a sliding window along the financial time series data. We then determine whether these constructed networks encapsulate observable properties, such as connectivity (given by the average degree) and transitivity (given by the clustering coefficient), of the time-varying complex network that underlies the market. We compute the time series of network characteristics of the generated networks, and compare it against known crisis periods. The analysis was done in the statistical software `R`, using the packages `igraph` [40], and `glmnet` [39]. The visualizations were generated by using the package `ggplot2` [41].

4.1 Data Description

Let us begin by describing the datasets used. Both datasets contain index data, which is used to compute log-returns, based on which extreme events are filtered. Using a sliding window, we generate a time varying network.

4.1.1 Dataset 1

Dataset 1 comprises of the time series of stock indices for 17 different economies, namely Argentina, Brazil, Chile, Colombia, Mexico, Peru, China, India, Indonesia, Korea, Malaysia, Philippines, Taiwan, Thailand, USA, and Japan. These MSCI indices were acquired from Datastream in US dollars and are recorded on a weekly basis. The period under study spans from January 1993 to December 2011, during which we have a total of 992 observations for each of the 16 time series. By using weekly data, we aim to minimize the potential issue of non-synchronous data to some extent. Weekly data also provides mitigation against spillover effects.

4.1.2 Dataset 2

A selection was made of closing time index data from 33 major stock markets across Asia, America, Europe, and Oceania, covering the period from May 3, 2004, to June 30, 2017, to construct dataset 2. The data was recorded on a daily basis. [13]

4.2 Network Characteristics using the Modified RRT

A time-varying network is generated using LASSO on the residual and recurrence times with the ℓ^1 penalty parameter set at $\lambda = 1$. This means that the optimization problem for the regression coefficients becomes:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{n-1}} \left\{ \frac{1}{\ell} \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right\|_2^2 + \left\| \beta \right\|_1 \right\}$$

The optimal choice of λ based on mean squared error was found to be in in the interval $(1.2, 1.6)$ in experiments, however, we choose $\lambda = 1$ to encourage more links being drawn per node as the optimal choice produced fewer edges than desired.

We compute the time series of four network characteristics of the time-varying complex network and heatmaps for indegree and outdegree for each node from data. We use average degree per node (Definition 2.0.4) and a new indicator, the mean signal strength per node,

based on the interpretation of the estimated coefficients in the regression as 'signal strengths' between two nodes. It is defined as:

Definition 4.2.1. ***Mean Signal Strength Per Node of a Network***
For a directed, weighted network, the mean signal strength is defined as the sum of all weights in the network. We can calculate this by summing over all the elements of the adjacency matrix and then dividing by the number of nodes: $\mathcal{A}$*:*

$$\bar{d}_{net} = \frac{\sum_{i,j=1}^{n} \mathcal{A}_{ij}}{n}$$

where n *is the total number of nodes and* e *is the total number of edges. The mean signal strength of a network can be interpreted as representing the net movement of extreme events in a network.*

Known crisis periods are highlighted in the graphs: **dark grey** for the 1997-1998 Asian Crisis, **light grey** for the DotCom Crash 2000-2002 and finally, **red** for the 2007-2008 US Subprime Crisis.

In this subsection, we consider the results of experiments on Dataset 1. We inspect the time series of three network characteristics and the behaviour of specific nodes via heatmaps, as described below:

1. Average degree per node (blue graphs), which is a measure of the connectivity of the network. (Definition 2.0.4)

2. Mean signal strength per node (green graphs), which is the sum of all entries of the adjacency matrix and represents the net movement of extreme events. (Definition 4.2.1)

3. Transitivity (pink graphs), which is a quantifier of the amount of clustering among nodes in the network, with the salient feature of highly connected nodes being weighted heavily as compared to nodes with low degree. (Definition 2.0.6)

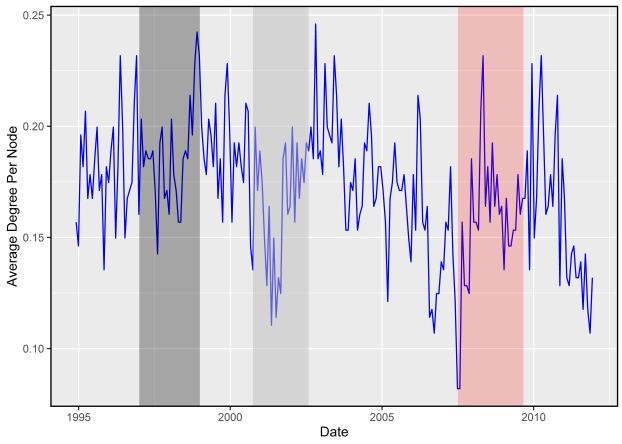
4. The **heatmaps** are useful in the identification of distress propagating sources and system stabilizing sinks. We present two heatmaps, one for indegree and the other

for outdegree. A high indegree or outdegree imply greater promience in spreading contagion.

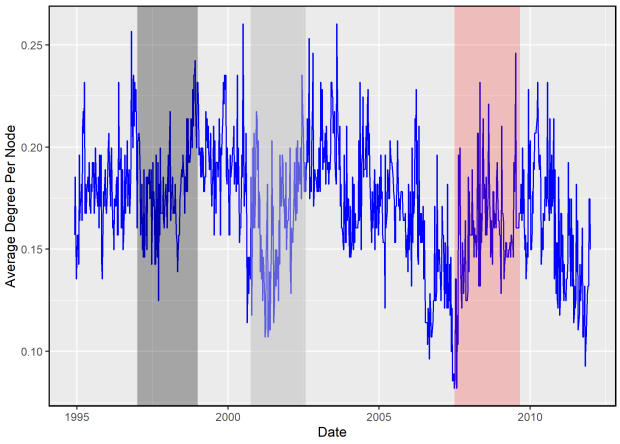
4.3 Effect of Varying Parameters

4.3.1 Effect of Step Size

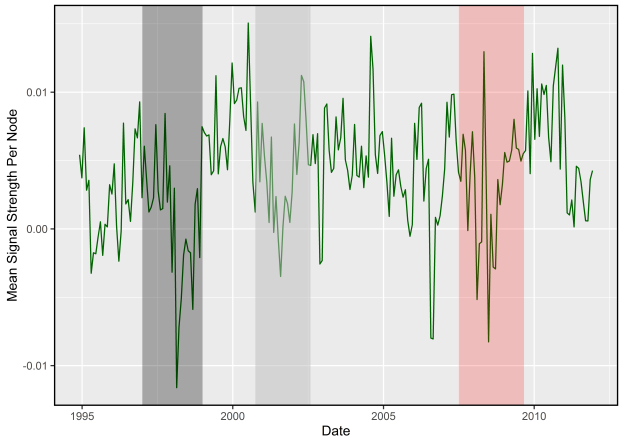
We observe that increasing the step size provides a smoothing effect on graph of the time series of network characteristics, however, it does not change the qualitative nature of the plot. This is illustrated in Figure 4.1 by contrasting various network characteristics computed with different step sizes.



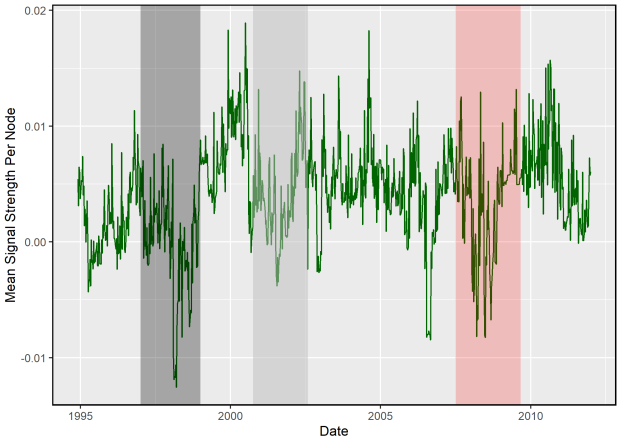
(a) Window: 100 weeks, Step: 4 weeks



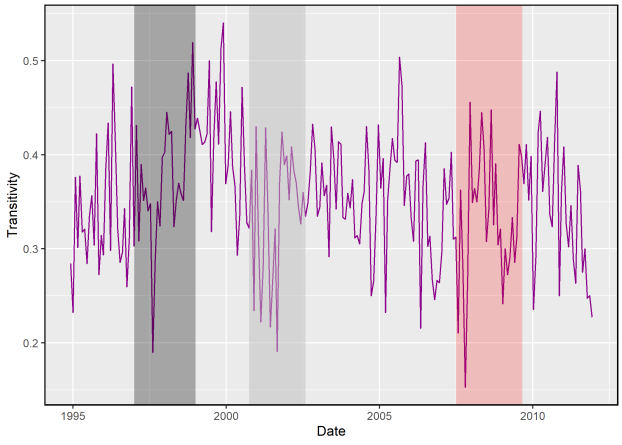
(b) Window: 100 weeks, Step: 1 week



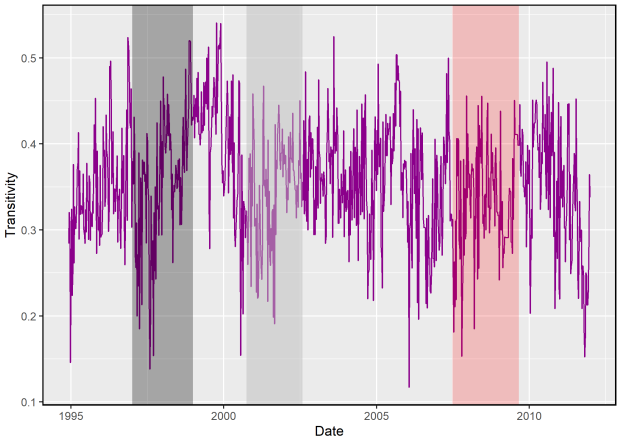
(c) Window: 100 weeks, Step: 4 weeks



(d) Window: 100 weeks, Step: 1 week



(e) Window: 100 weeks, Step: 4 weeks

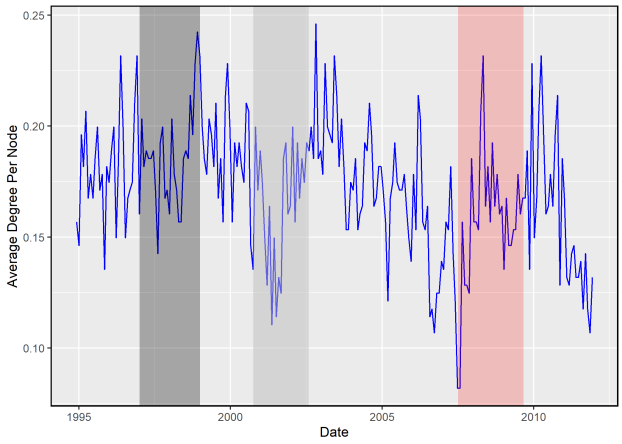


(f) Window: 100 weeks, Step: 1 week

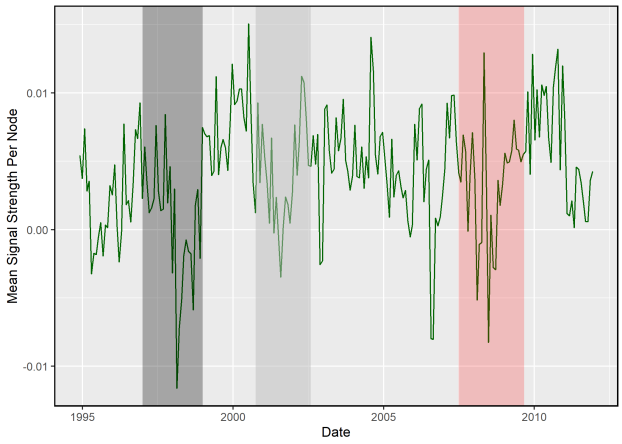
Figure 4.1: Effect of changing step size for selected network characteristics, from Dataset 1. All graphs are for the bottom 15 %ile of points and the penalty parameter has been set to $\lambda = 1$.

4.3.2 Effect of Window Length

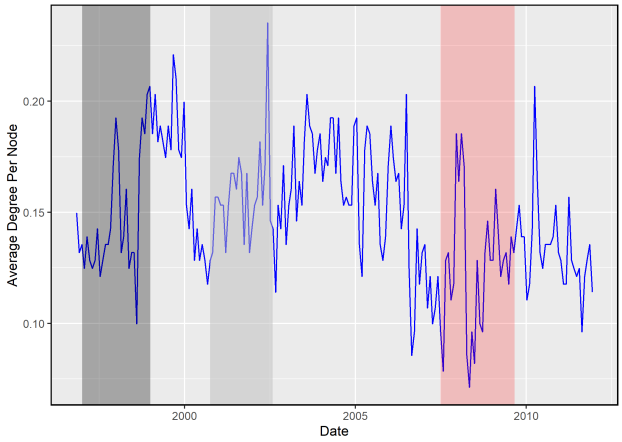
Figure 4.2 showcases the effect of changing step sizes for average degree and the mean signal strength in Dataset 1. We can observe a trade-off effect while changing the size of window length. A larger window means more data for the model to crunch, but the amount of data relevant to contagion is lesser. On the other hand, a smaller window will give the model data that is concentrated around events of interest, but consequently, there is lesser data that actually captures the event.



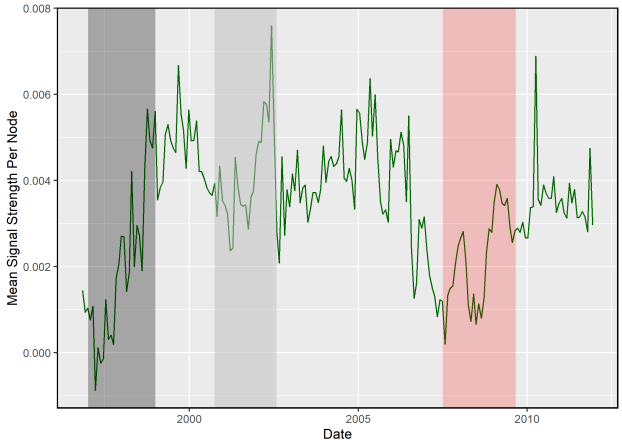
(a) Window: 100 weeks, Step: 4 week



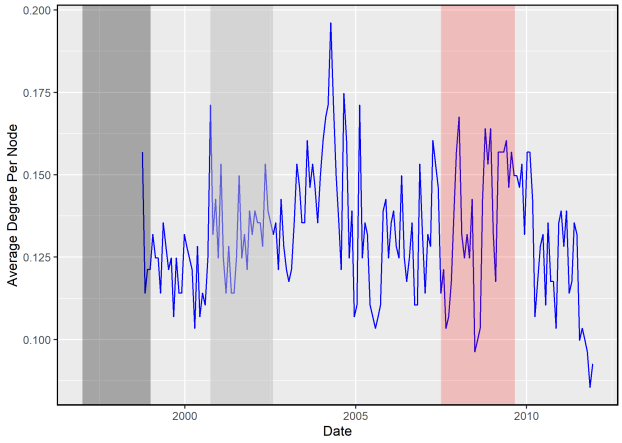
(b) Window: 100 weeks, Step: 4 weeks



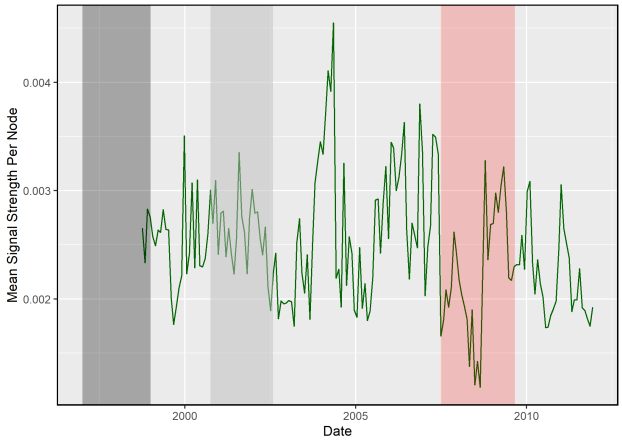
(c) Window: 200 weeks, Step: 4 weeks



(d) Window: 200 weeks, Step: 4 weeks



(e) Window: 300 weeks, Step: 4 weeks



(f) Window: 300 weeks, Step: 4 weeks

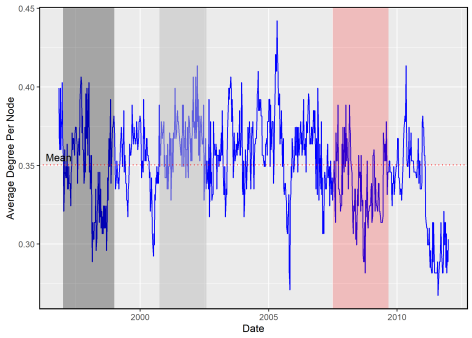
Figure 4.2: Effect of changing the window on selected network characteristics, for the bottom 15 %ile of points from Dataset 1 with a step of 4 weeks with penalty $\lambda = 1$.

4.3.3 Effect of Changing Penalty Parameter λ

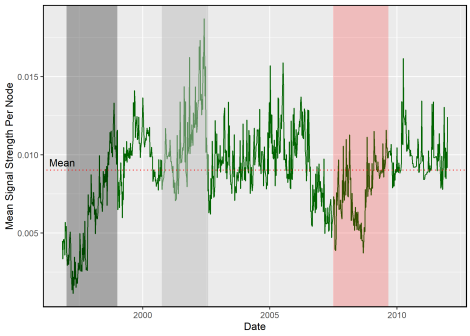
For $\lambda = 0$, the LASSO reduces to Ordinary Least Squares (OLS) estimation, and hence, the generated networks have $n^2 - n$ edges for n nodes. The value of λ can be used to force sparsity in the network: higher the penalty, higher the sparsity. For values of λ which are significantly greater than the optimal value, the shrinkage of the coefficient vector is extreme enough to force the network to have zero nodes.

We can observe this progressive effect on the average degree and the mean signal strength for Dataset 1 in Figure 4.3. As the value of λ increases, the mean value of the network characteristics over the dataset fall. In Fig 4.3g, we notice that the average degree per node goes down to zero for multiple time windows.

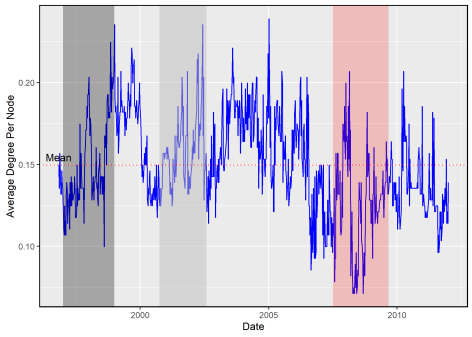
Hence, there is a trade-off effect in play here as well. Larger values of the penalty parameter generated networks with no significant relationships, but values that are too small lead to spurious correlations creeping into the estimates due to random error and produce a dense network.



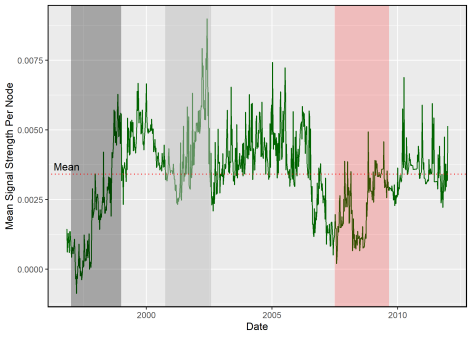
(a) $\lambda = 0.5$



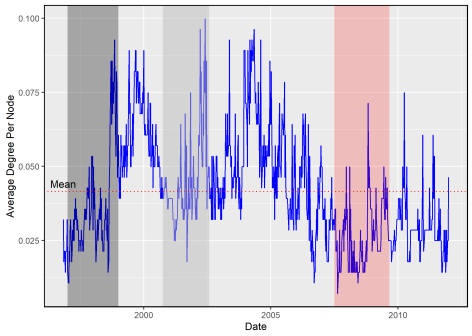
(b) $\lambda = 0.5$



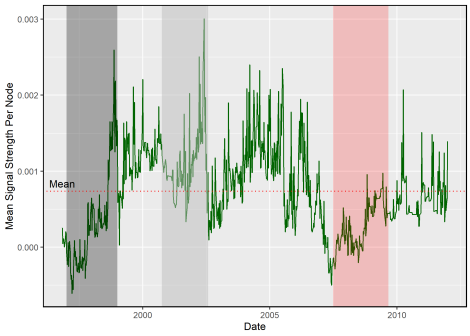
(c) $\lambda = 1$



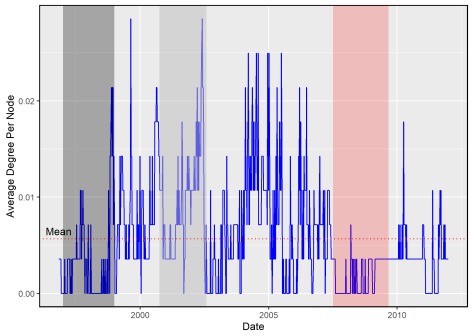
(d) $\lambda = 1$



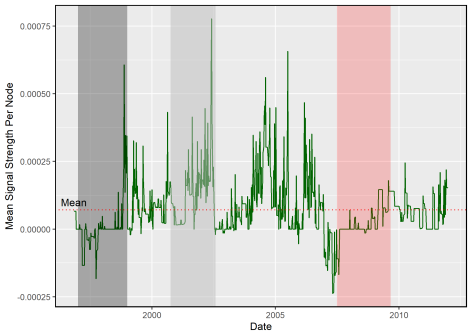
(e) $\lambda = 1.5$



(f) $\lambda = 1.5$



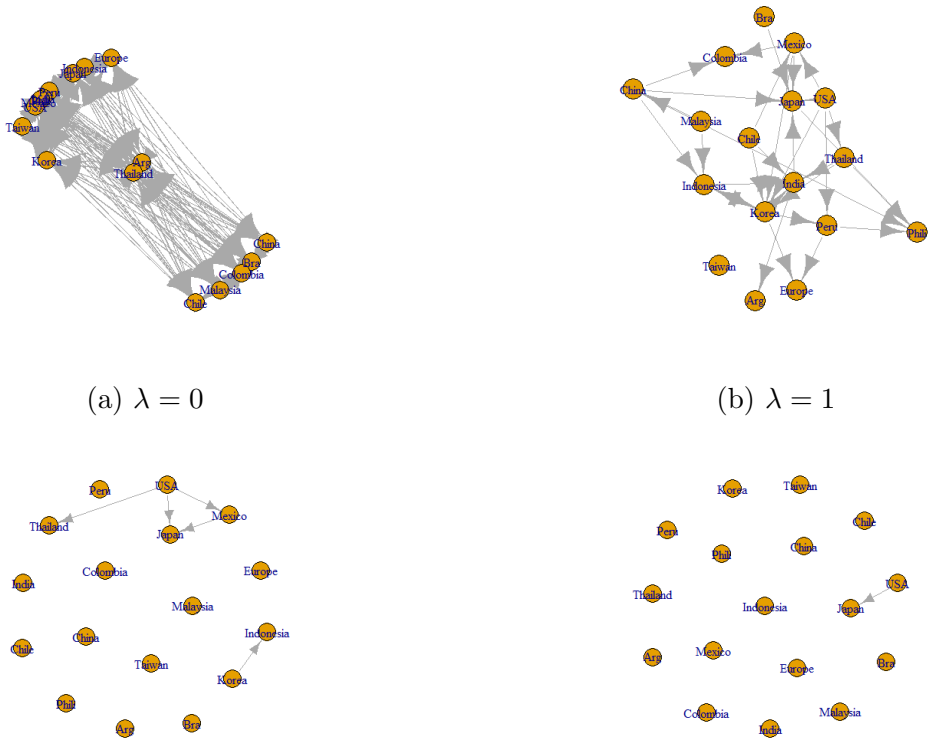
(g) $\lambda = 2$



(h) $\lambda = 2$

Figure 4.3: Effect of changing the penalty parameter λ on selected network characteristics, for the bottom 15 %ile of points from Dataset 1 with a window of 200 weeks and a step of 1 week.

We now present a visual illustration of the same effect in Figure 4.4. We consider the bottom 15 %ile of points in Dataset 1, taking a window of 200 weeks and a step of 4 weeks (which is approximately a month), and draw the network for the last iteration of the algorithm, corresponding to the time window around 2011. The sparsity in the generated networks increases with the value of λ, until finally, in Fig 4.4d, only one links is drawn in the network.



(a) $\lambda = 0$ (b) $\lambda = 1$

(c) $\lambda = 1.5$ (d) $\lambda = 2$

Figure 4.4: Network representation of the effect of using LASSO for coefficient estimation on Dataset 1, with a window of 200 weeks and a step of 4 weeks, for the bottom 15 %ile of points.

4.3.4 Optimal Choice of Parameters

From the various numerical experiments that were undertaken, the optimal values of the parameters were found to be as follows:

- **Window Length** The ideal window length for both datasets was found to be 200 units. A smaller window size leads to fewer data points for the estimation, whereas a larger size glosses over events of interest.

- **Step Size** The step size does not have a sizeable impact on the qualitative nature of the generated plots. We choose the step size to be close to a month (4 weeks for Dataset 1 and 20 days for Dataset 2).

- **Penalty Parameter** For the penalty parameter, cross-validation showed that for most experiments, the lowest mean squared error was achieved at values around 1. Hence, we choose $\lambda = 1$ as the standard penalty parameter for all experiments.

4.4 Results for Dataset 1

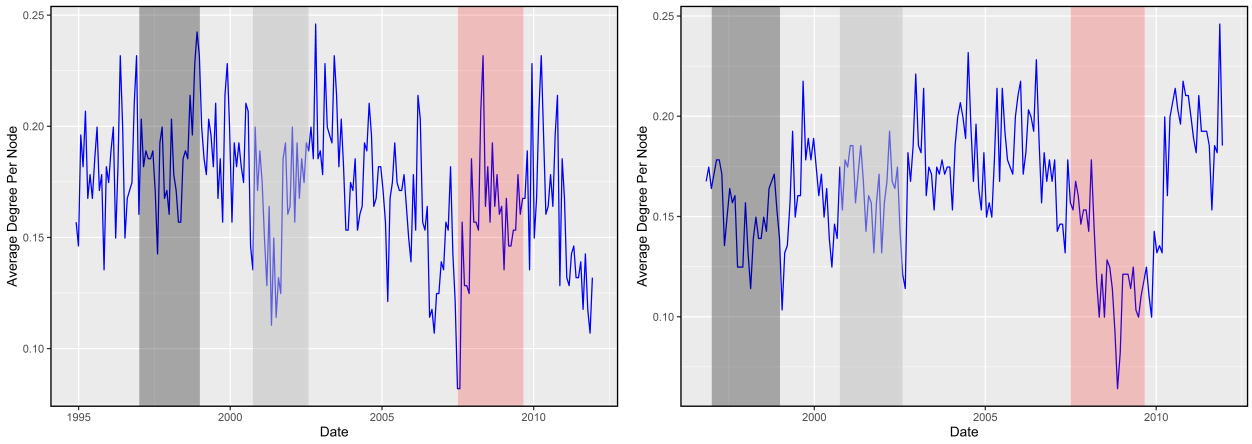
4.4.1 Network Characteristics

Average Degree

Figure 4.5 showcases the average degree per node for Dataset 1. We choose a step size of 4 weeks, which approximately corresponds to a month in real time. In Figure 4.6, we set the window to 200 weeks and the step to 4 weeks, and vary the percentile threshold for selecting extreme events.

We can observe that this graph shows varying behaviour across all three marked crisis periods. One consistent feature is that the plot reaches a minima in or very nearly around the red region, which represents the 2008 US Subpriime Crisis. In both Figure 4.5a and Figure 4.5b, the light grey area shows a prominent dip in the value of connectivity.

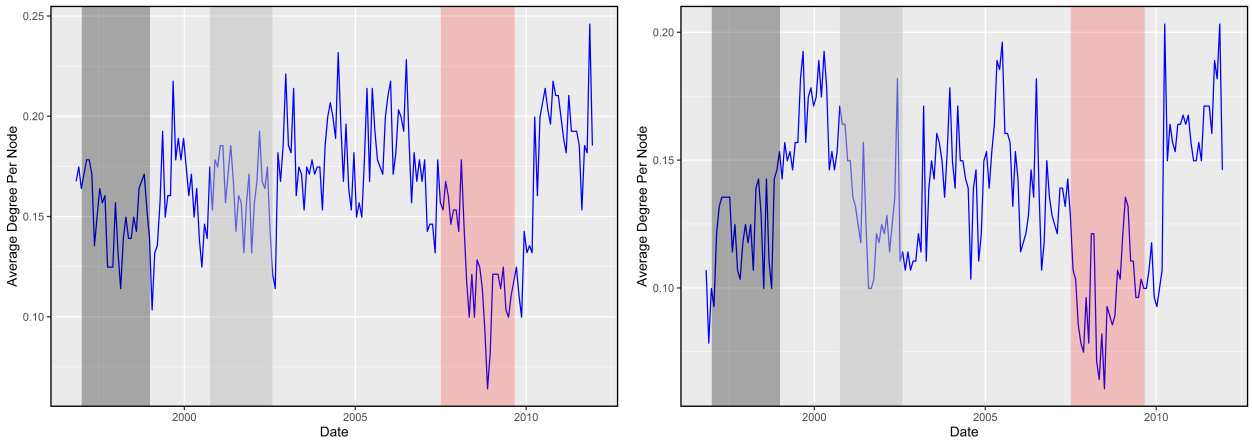
In Figure 4.6, we can observe the effect of changing the percentile threshold. The presence of the minimum in the red region is preserved, and, the qualitative features remain mostly similar. We do notice that the value of the absolute minimum shrinks with an increase in the percentile value.



(a) Window: 100 weeks, Step: 4 weeks

(b) Window: 200 weeks, Step: 4 weeks

Figure 4.5: Average Degree per node for Dataset 1 for the bottom 15 %ile of points, over a step size of 4 weeks and different window lengths.



(a) Threshold: 15 %ile

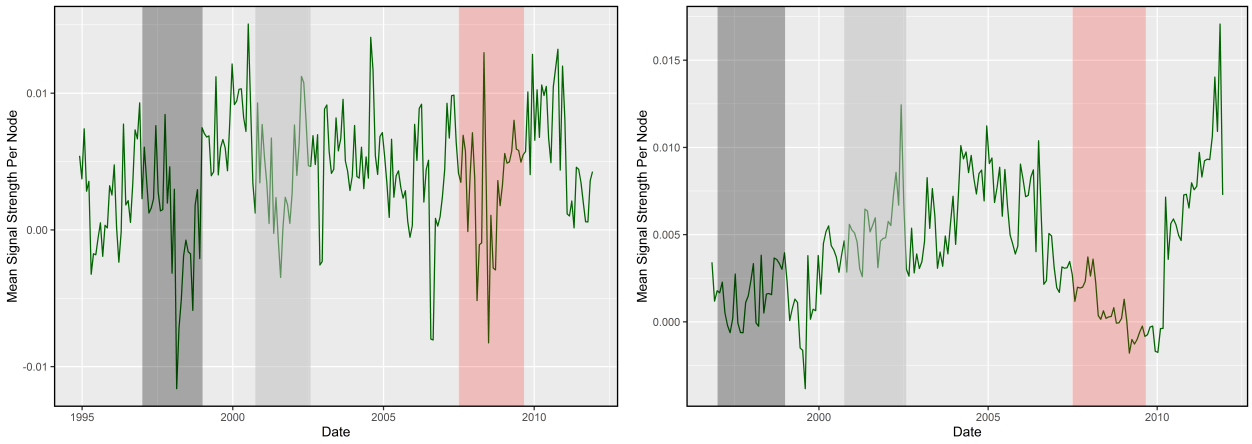
(b) Threshold: 10 %ile

Figure 4.6: Average Degree per node for Dataset 1 for a window of 200 weeks and step of 4 weeks, over different percentile values for the threshold of extreme events.

Mean Signal Strength

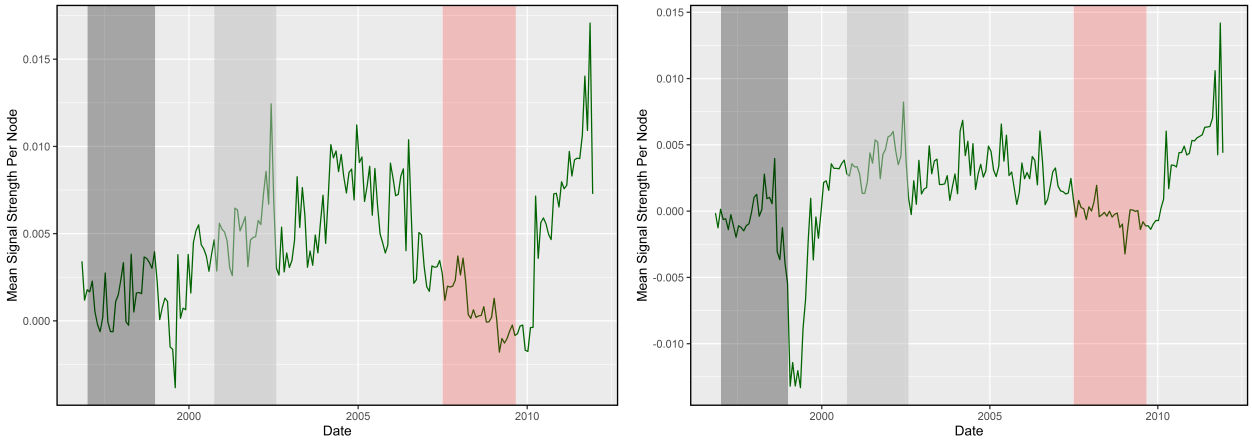
The mean signal strength, obtained by summing over all entries of the adjacency matrix of a generated network, gives a sense of the direction of net movement in the network. Figure 4.7 gives a sense shows the time series of mean signal strength per node for two experimental configurations. With a resolution of 100 weeks, we can see that the plot shows prominent dips around crisis period in Figure 4.7a, but the same cannot be observed in 4.7b, possiibly due to the larger window size.

Figure 4.8 shows the behaviour of meann signal strength as we increae the percentile threshold. The dipping behaviour of the plots, specifically in the red region, is more pronounced as we increase the percentile value. This could be due to more precise estimation of links when we select a larger number of points for the regression by increasing the threshold.



(a) Window: 100 weeks, Step: 4 weeks (b) Window: 200 weeks, Step: 4 weeks

Figure 4.7: Mean Signal Strength per node for Dataset 1 for the bottom 15 %ile of points, over a step size of 4 weeks and different window lengths.



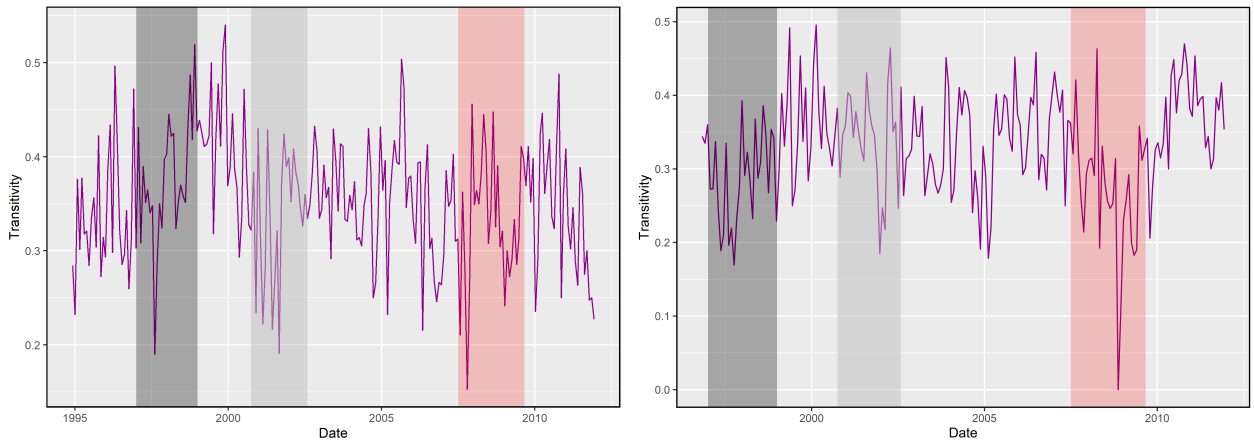
(a) Threshold: 15 %ile (b) Threshold: 10 %ile

Figure 4.8: Mean Signal Strength per node for Dataset 1 for a window of 200 weeks and step of 4 weeks, over different percentile values for the threshold of extreme events.

Transitivity

The behaviour of the time series transitivity exhibits high variance, but maintains the dipping feature seenn in the previous network characteristics. In Figure 4.9a, the plot reaches local minima during crisis periods. A sharp, prominent dip can be seen in the red area in Figure 4.9b.

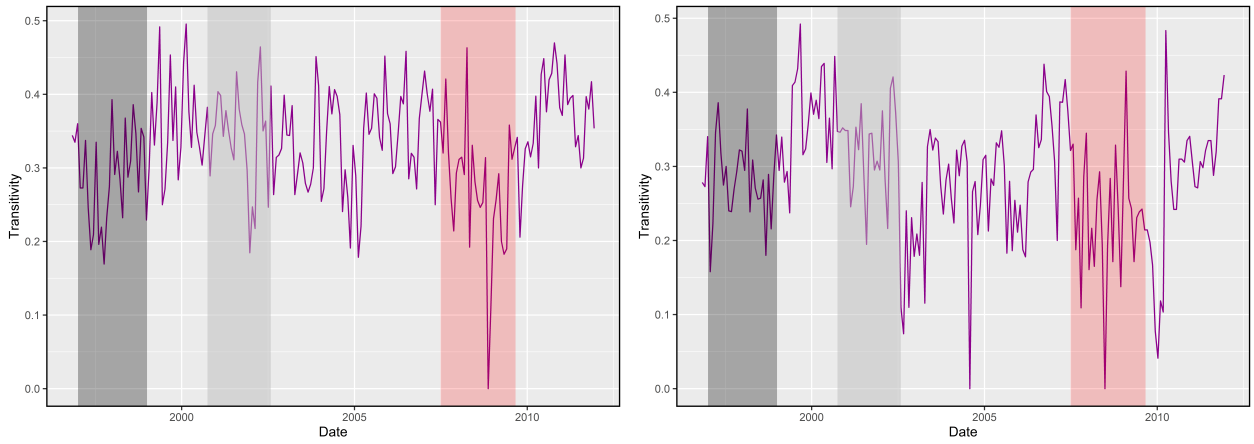
The effect of changing percentile values (Fig 4.10) is less clear in the case of transitivity. This may be due to the large window length of 300 weeks, which correponds to 6 years.



(a) Window: 100 weeks, Step: 4 weeks

(b) Window: 200 weeks, Step: 4 weeks

Figure 4.9: Transitivity for Dataset 1 for the bottom 15 %ile of points, over a step size of 4 weeks and different window lengths.



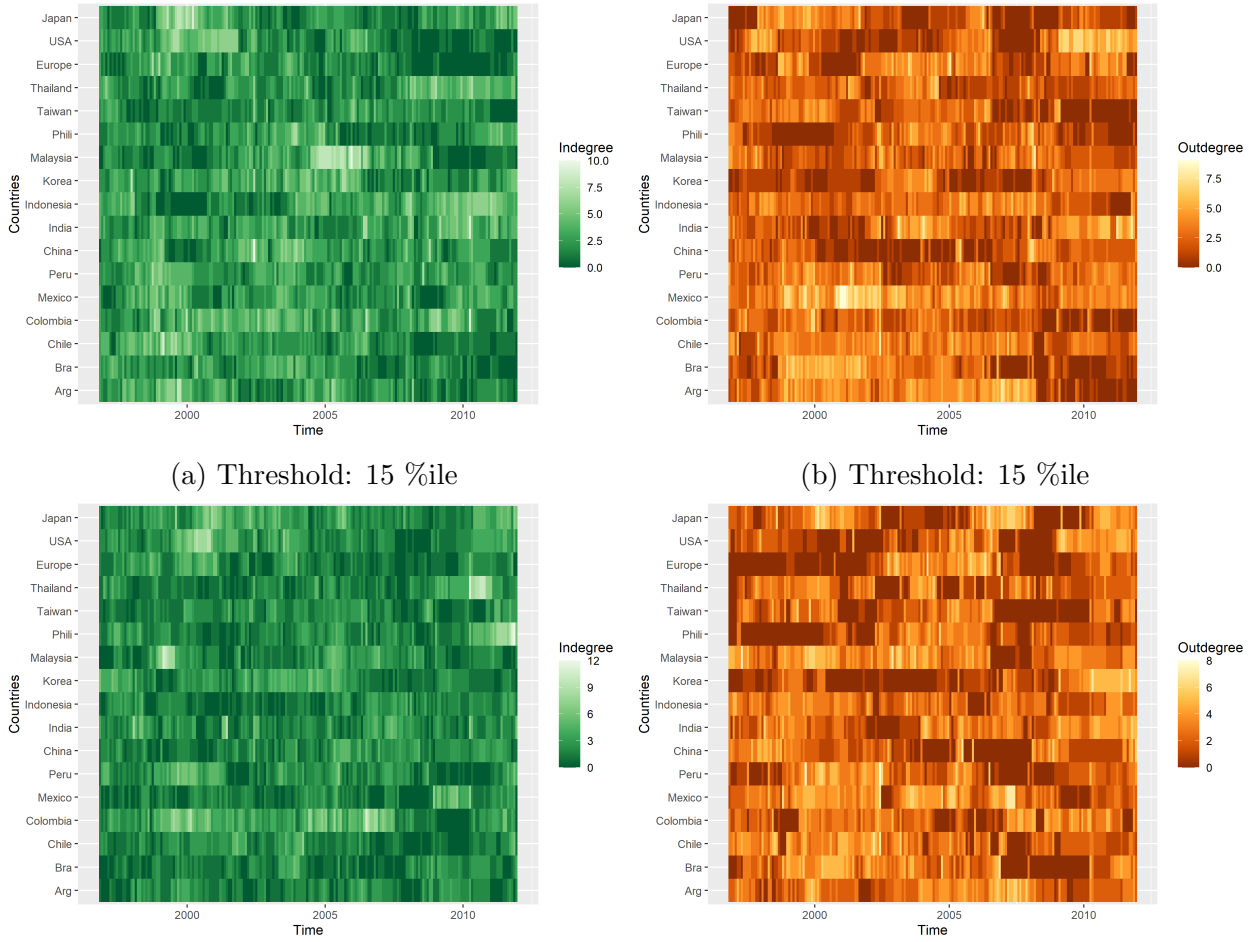
(a) Threshold: 5 %ile

(b) Threshold: 10 %ile

Figure 4.10: Transitivity for Dataset 1 for a window of 200 weeks and step of 4 weeks, over different percentile values for the threshold of extreme events.

Heatmaps

Heatmaps are useful for simultaneously visualising the time series of specific properties for all nodes in a network. Figure 4.11 depicts the time evolution of indegree and outdegree for the nodes of the generated network. Darker patches imply a low value of indegree/outdegree.



(a) Threshold: 15 %ile

(b) Threshold: 15 %ile

(c) Threshold: 10 %ile

(d) Threshold: 10 %ile

Figure 4.11: Heatmaps of Indegree and Outdegree for Dataset 1, corresponding to the experiments in Fig 4.6. We use a window of 200 weeks and a step of 4 weeks.

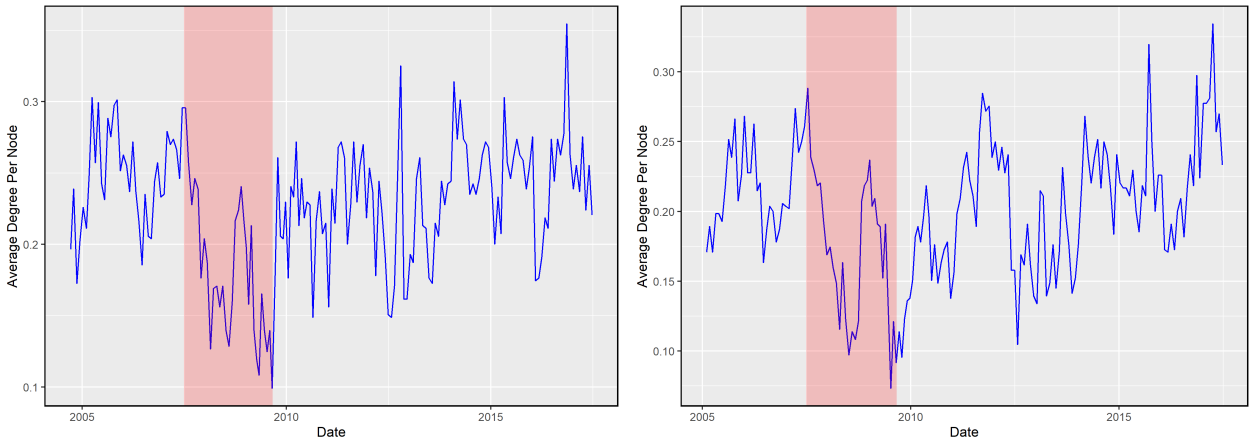
4.5 Results for Dataset 2

4.5.1 Network Characteristics

Average Degree

Figure 4.12 shows the average degree per node evolving in time in Dataset 2. We see that the graph reaches its minimum during the highlighted crisis period. Furthermore, within the red region, sandiwched between two dips is a steep increase in the connectivity of the generated networks.

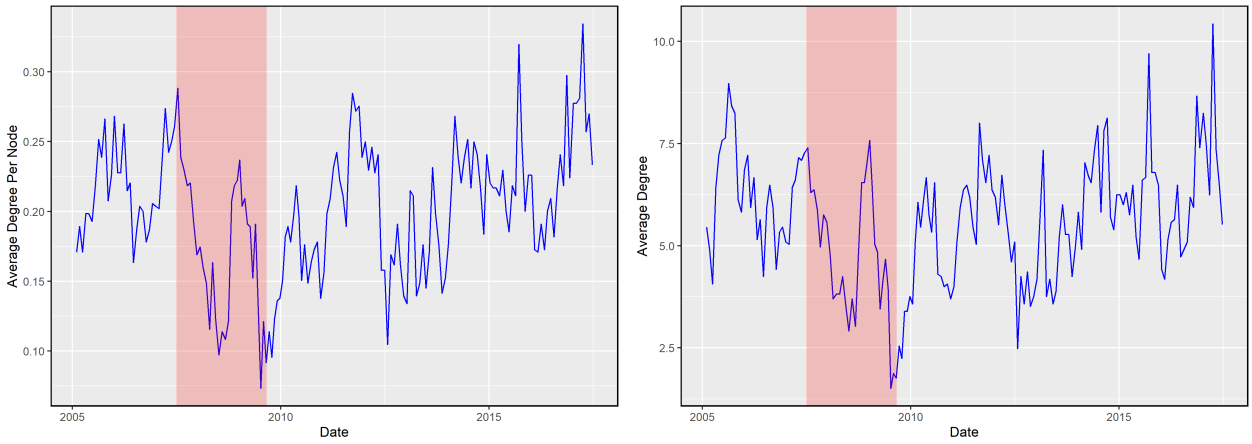
In Figure 4.13, we see the effect of changing the percentile threshold on the average degree. Since the data is recorded daily, it is highly resolved and this provides an ample number of points for estimation of links. Due to this, the qualititive features of the graphs do not change over the panel. We still observe deep dips around the crisis period, along with a transient peak within the crisis period.



(a) Window: 100 days, Step: 20 days

(b) Window: 200 days, Step: 20 days

Figure 4.12: Average Degree per node for Dataset 2 for the bottom 15 %ile of points, over a step size of 20 days and different window lengths.



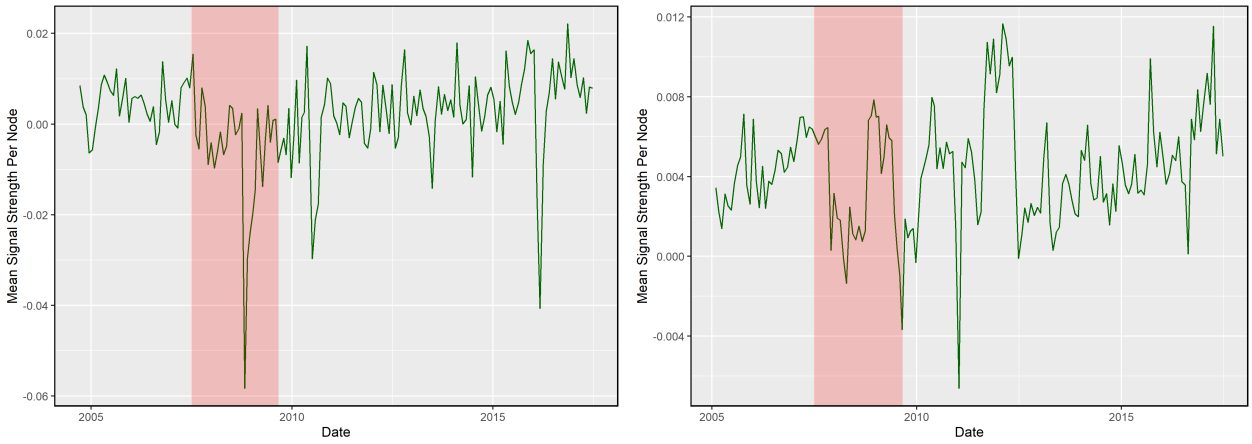
(a) Threshold: 15 %ile

(b) Threshold: 10 %ile

Figure 4.13: Average Degree per node for Dataset 2 for a window of 200 days and step of 20 days, over different percentile values for the bottom threshold of extreme events.

Mean Signal Strength

The behaviour of mean signal strength on Dataset 2 is especially interesting. The highlighted crisis period shows a sharp, prominent dip in Figures 4.14a and 4.15. The crisis period is easily discernible compared to the plots for average degree. This implies that the mean signal strength displays more information pertaining to the behaviour of the entire system and hence, is a better marker for systemic contagion as compared to average degree.

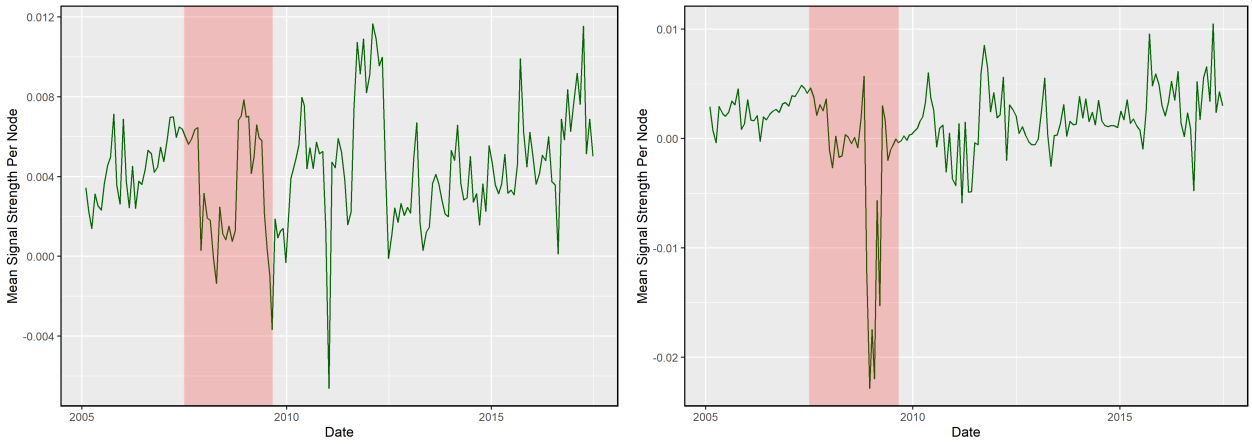


(a) Window: 100 days, Step: 20 days

(b) Window: 200 days, Step: 20 days

Figure 4.14: Mean Signal Strength per node for Dataset 2 for the bottom 15 %ile of points, over a step size of 20 days and different window lengths.

Wehn we look at the effect of changing the percentile values for mean signal strength in Figure 4.15, we observe that all the plots exhibit a sharp dip during the crisis period, but as the percentile threshold increases, more variation is seen in the plots. In fact, Figure **??** resembled the plot for average degree. This may be due to 0.15 being too low a barrier to successfully capture extreme events.



(a) Threshold: 15 %ile

(b) Threshold: 10 %ile

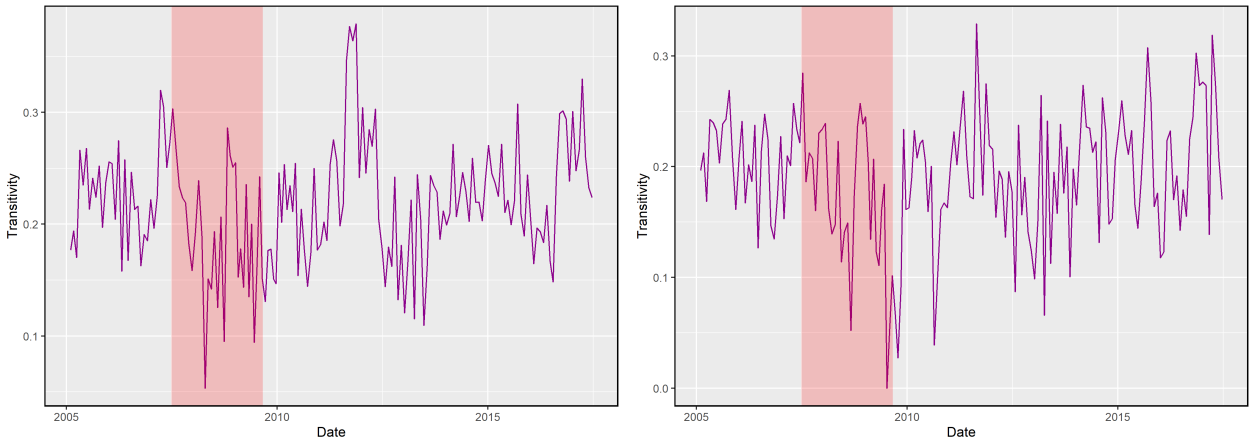
Figure 4.15: Mean Signal Strength per node for Dataset 2 for a window of 200 days and step of 20 days, over different percentile values for the bottom threshold of extreme events.

Transitivity

Once again, we observe that transitivity shows higher variance compared to the other network indicators. Similar to the case with Dataset 1, the dips are located around the highlighted crisis period.

(a) Window: 100 days, Step: 20 days (b) Window: 200 days, Step: 20 days

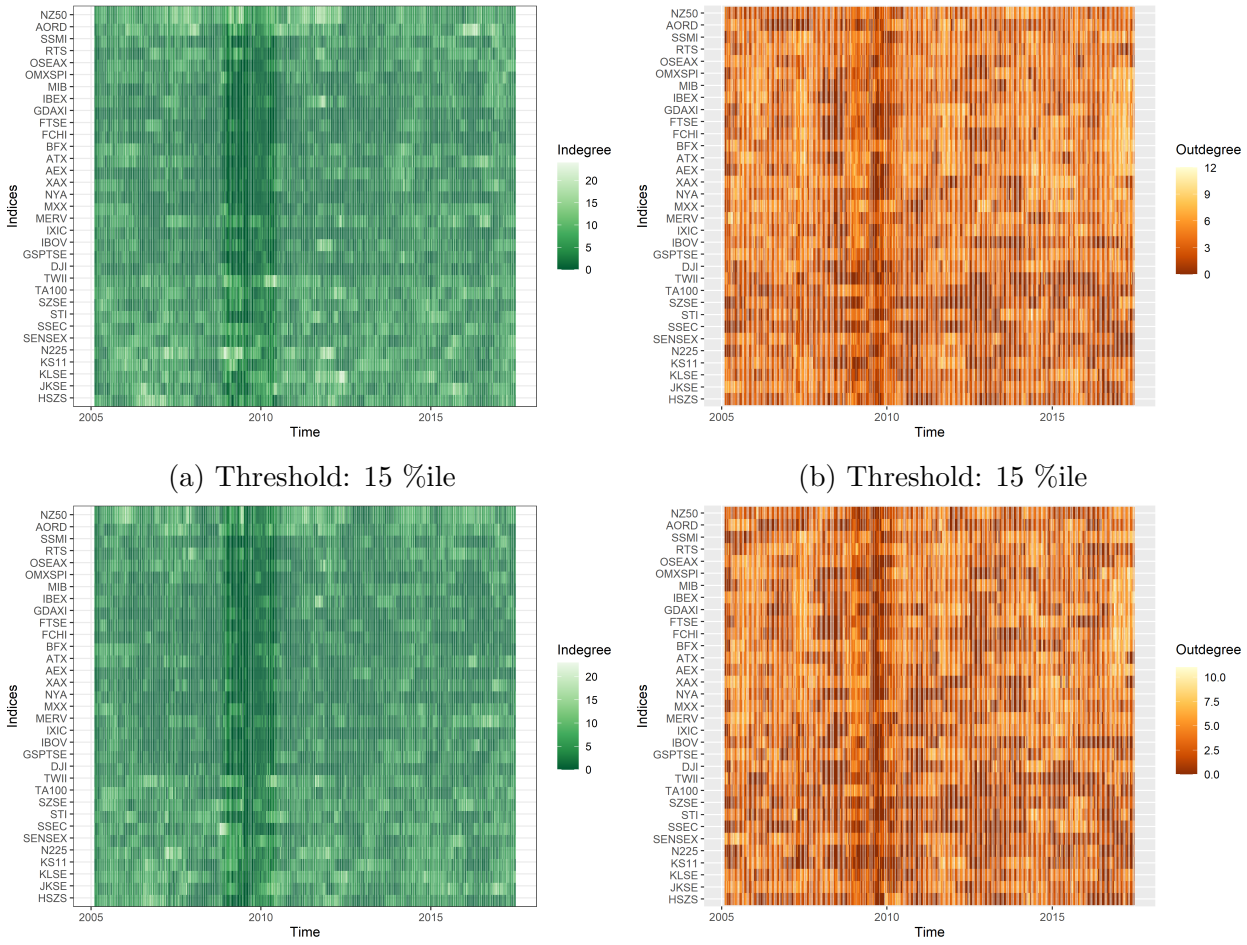
Figure 4.16: Transitivity for Dataset 2 for the bottom 15 %ile of points, over a step size of 20 days and different window lengths.



(a) Threshold: 15 %ile (b) Threshold: 10 %ile

Figure 4.17: Transitivity for Dataset 2 for a window of 200 days and step of 20 days, over different percentile values for the bottom threshold of extreme events.

Heatmaps



(a) Threshold: 15 %ile

(b) Threshold: 15 %ile

(c) Threshold: 10 %ile

(d) Threshold: 10 %ile

Figure 4.18: Heatmaps of Indegree and Outdegree for Dataset 2, corresponding to the experiments in Fig 4.13. We use a window of 200 days and a step of 5 days.

Chapter 5

Conclusion

In this thesis, we developed a new method for detection and quantification of extreme dependence in financial markets. Using this model, we were able to generate a network representation of the market system from two datasets depicting the global financial market. We then compute summary statistics for the network, as these are proxies for the state of the system. In particular, we analyse the time series of average degree and net weighted degree for the data. We attempt to validate the method by using known crisis periods as a benchmark and assessing if these summary statistics show anomalous behaviour during these times.

The selected network indicators exhibited deep dips during crisis periods, especially in mean signal strength. This appears to be contrary to the interpretation of the estimated coefficients, as we expect the coefficients to increase in numerical value during crisis periods. This behaviour was observed in the connectivity of the graphs as well.

A plausible interpretation that reconciles the observations with the theory could be related to the fact that the model presented in this thesis looks specifically at the *lower tail dependence*. While the net correlation goes up during crisis periods, the lower tail dependence might not necessarily go up as well. In terms of macroeconomics, these observations may imply that extreme events on the lower side of the distribution cease moving together during the middle of the crisis period, but resume this behaviour as the crisis starts to fade, leading to an initial dip and then a rise in the network indicators.

The mean signal strength was found to be a better indicator of contagion as compared to the average degree and the transitivity of the network consistently across different experiments. A possible interpretation of this trend is that the mean signal strength encapsulates relevant information about the state of the system and is thus a better marker of financial contagion.

Future directions for this work could include but aren't limited to:

1. Modelling with fewer assumptions and looking for extensions beyond the domain of generalized linear models.
2. Using more powerful statistical estimation tools like Elastic Net instead of LASSO to better tackle the problem of high dimensionality.
3. Using volatility series as opposed to index data to detect the transmission of extreme volatility and study volatility clustering.
4. Constructing a framework for determining market health and assessing system risk using network characteristics.

The author would like to conclude this thesis with a fairly common, but maybe the most important aphorism in statistics:

All models are wrong, but some are useful. ■■■

Bibliography

[1] Pavel Tsankov. OVERVIEW OF NETWORK-BASED METHODS FOR ANALYZING FINANCIAL MARKETS. *Proceedings of the Technical University of Sofia*, 71(1):1–7, March 2021.

[2] Kausik Chaudhuri, Rituparna Sen, and Zheng Tan. Testing extreme dependence in financial time series. *Economic Modelling*, 73:378–394, June 2018.

[3] Sumanta Basu. A System-Wide Approach to Measure Connectivity in the Financial Sector. *SSRN Electronic Journal*, 2016.

[4] Mervyn A King and Sushil Wadhwani. Transmission of volatility between stock markets. *The Review of Financial Studies*, 3(1):5–33, 1990.

[5] Sang Bin Lee and Kwang Jung Kim. DOES THE OCTOBER 1987 CRASH STRENGTHEN THE CO-MOVEMENTS AMONG NATIONAL STOCK MARKETS? *Review of Financial Economics*, 3(1):89–102, September 1993.

[6] Dominik Blatt, Bertrand Candelon, and Hans Manner. Detecting contagion in a multivariate time series system: An application to sovereign bond markets in europe. *Journal of Banking and Finance*, 59:1–13, 2015.

[7] Rosario N Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11:193–197, 1999.

[8] Giovanni Bonanno, Guido Caldarelli, Fabrizio Lillo, Salvatore Micciche, Nicolas Vandewalle, and Rosario Nunzio Mantegna. Networks of equities in financial markets. *The European Physical Journal B*, 38:363–371, 2004.

[9] Tomáš Vỳrost, Štefan Lyócsa, and Eduard Baumöhl. Granger causality stock market networks: Temporal proximity and preferential attachment. *Physica A: Statistical Mechanics and its Applications*, 427:262–276, 2015.

[10] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[11] Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559, 2012.

[12] Mert Demirer, Francis X. Diebold, Laura Liu, and Kamil Yilmaz. Estimating global bank network connectedness. *Journal of Applied Econometrics*, 33(1):1–15, January 2018.

[13] Qiuhong Zheng and Liangrong Song. Dynamic Contagion of Systemic Risks on Global Main Equity Markets Based on Granger Causality Networks. *Discrete Dynamics in Nature and Society*, 2018:1–13, August 2018.

[14] Mariana Durcheva and Pavel Tsankov. Granger causality networks of S&P 500 stocks. page 110014, Tomsk, Russia, 2021.

[15] Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

[16] Takayuki Mizuno, Hideki Takayasu, and Misako Takayasu. Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications*, 364:336–342, 2006.

[17] Boris Podobnik, Duan Wang, Davor Horvatic, Ivo Grosse, and H Eugene Stanley. Time-lag cross-correlations in collective phenomena. *Europhysics Letters*, 90(6):68001, 2010.

[18] Adrian Tobias and Markus K Brunnermeier. Covar. *The American Economic Review*, 106(7):1705, 2016.

[19] Giovanni Covi, Mehmet Ziya Gorpe, and Christoffer Kok. Comap: mapping contagion in the euro area banking sector. *Journal of Financial Stability*, 53:100814, 2021.

[20] Viral Acharya, Robert Engle, and Matthew Richardson. Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64, 2012.

[21] Leo Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37(4):373–384, 1995.

[22] Marta Bańbura, Domenico Giannone, and Lucrezia Reichlin. Large bayesian vector auto regressions. *Journal of applied Econometrics*, 25(1):71–92, 2010.

[23] Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. 2015.

[24] Cedric E. Ginestet, Jun Li, Prakash Balachandran, Steven Rosenberg, and Eric D. Kolaczyk. Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, 11(2), June 2017.

[25] Matteo Barigozzi and Marc Hallin. A network analysis of the volatility of high dimensional financial series. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, pages 581–605, 2017.

[26] Stefano Giglio, Bryan Kelly, and Seth Pruitt. Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics*, 119(3):457–471, 2016.

[27] Linda Allen, Turan G Bali, and Yi Tang. Does systemic risk in the financial sector predict future economic downturns? *The Review of Financial Studies*, 25(10):3000–3036, 2012.

[28] Fenghua Wen, Xin Yang, and Wei-Xing Zhou. Tail dependence networks of global stock markets. *International Journal of Finance & Economics*, 24(1):558–567, January 2019.

[29] Xiurong Chen, Aimin Hao, and Yali Li. The impact of financial contagion on real economy-An empirical research based on combination of complex network technology and spatial econometrics model. *PLOS ONE*, 15(3):e0229913, March 2020.

[30] Xiangyun Gao, Shupei Huang, Xiaoqi Sun, Xiaoqing Hao, and Feng An. Modelling cointegration and Granger causality network to detect long-term equilibrium and diffusion paths in the financial system. *Royal Society Open Science*, 5(3):172092, March 2018.

[31] Christian Brownlees, Eulàlia Nualart, and Yucheng Sun. Realized networks. *Journal of Applied Econometrics*, 33(7):986–1006, November 2018.

[32] Daniele Bianchi, Monica Billio, Roberto Casarin, and Massimo Guidolin. Modeling contagion and systemic risk. *Available at SSRN*, 2015.

[33] Adam Ashcraft. *MBS ratings and the mortgage credit boom*. DIANE Publishing, 2010.

[34] Michael Gofman. Efficiency and stability of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124(1):113–146, 2017.

[35] Morten L Bech and Enghin Atalay. The topology of the federal funds market. *Physica A: Statistical mechanics and its applications*, 389(22):5223–5246, 2010.

[36] Ben Craig and Goetz Von Peter. Interbank tiering and money center banks. *Journal of Financial Intermediation*, 23(3):322–347, 2014.

[37] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

[38] Peter McCullagh. *Generalized linear models*. Routledge, 2019.

[39] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent, 2010.

[40] Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.

[41] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.