Journal of Physics: Complexity

OPEN ACCESS

PAPER

Quantifying impact and response in markets using information filtering networks

Isobel Seabrook[1,2], Fabio Caccioli[1,3] and Tomaso Aste[1,3,*]

[1] Department of Computer Science, University College London, Gower Street, WC1E 6EA London, United Kingdom
[2] Consumer Investments, Financial Conduct Authority, London E20 1JN, United Kingdom
[3] Systemic Risk Centre, London School of Economics, London, United Kingdom
[*] Author to whom any correspondence should be addressed.

E-mail: t.aste@ucl.ac.uk

Keywords: stress testing, systemic risk, elliptical conditional probability, financial modeling

RECEIVED
27 April 2021

REVISED
9 April 2022

ACCEPTED FOR PUBLICATION
13 April 2022

PUBLISHED
9 May 2022

Abstract

We present a novel methodology to quantify the 'impact' of and 'response' to market shocks. We apply shocks to a group of stocks in a part of the market, and we quantify the effects in terms of average losses on another part of the market using a sparse probabilistic elliptical model for the multivariate return distribution of the whole market. Sparsity is introduced with an L_0-norm regularization, which forces to zero some elements of the inverse covariance according to a dependency structure inferred from an information filtering network. Our study concerns the FTSE 100 and 250 markets and analyzes impact and response to shocks both applied to and received from individual stocks and group of stocks. We observe that the shock pattern is related to the structure of the network associated with the sparse structure of the inverse covariance of stock log-returns. Central sectors appear more likely to be affected by shocks, and stocks with a large level of underlying diversification have a larger impact on the rest of the market when experiencing shocks. By analyzing the system during times of crisis and comparative market calmness, we observe changes in the shock patterns with a convergent behavior in times of crisis.

1. Introduction

Stress testing allows us to understand how a system reacts to external shocks. For a complex system, which is composed of many interacting units, this requires an understanding of how shocks propagate between different units, and how these units depend on each other.

When the relationships between the constituents of the system are explicitly known (for instance, we may have knowledge of the trading relations between firms in a supply chain, or of mutual exposures in an interbank network), one can construct a mechanistic model to probe the response of the system to shocks. This approach has been applied to the analysis of diverse systems such as financial networks [1, 2], power grids [3, 4], the internet [5, 6], foodwebs [7, 8].

When we have no explicit knowledge of the relations between different constituents, these dependencies need instead to be inferred from multivariate time series that specify how the state of each unit changes over time. In this context, network information filtering techniques proved useful in a variety of domains ranging from the study of financial markets [9], to biology [10], semantics [11], psychology and the human brain [12].

In this paper, we focus on the stock market, and we develop a stress testing framework based on network information filtering to quantify the propagation of shocks between stocks and the reaction of the market to exogenous shocks.

Using this framework, we can identify stocks that either have the largest impact on other stocks, or show the largest response to shocks from others.

We measure stress propagation in terms of the conditional distribution of losses, which we compute by modeling the return distributions of the entire market in terms of a multivariate elliptical probability

distribution [13]. An underlying assumption of our analysis is that prices (and therefore returns) contain all available information about the value of stocks.

Impact is quantified as the average losses caused by a set of stressed stocks on the rest of the system. Conversely, response is quantified as the average losses suffered by a set of stocks when the rest of the system is stressed. We apply this method to different states of the FTSE 100 and 250 markets.

The estimation of the multivariate probability is made accurate by using sparse inverse covariance estimation, where the non-zero elements are the edges of an information filtering network. We observe that the structure of this network is related to the behavior of the system when stressed. In particular, we observe that impact and response of individual nodes are both related to their centrality, but when we consider groups of nodes, the most central groups have a higher response but lower impact due to the large internal effect that the related stocks have on each other.

With a regression analysis, we investigate if the impact and response measures of industry supersectors are affected by their centrality, and we find that the size and fraction of links shared by the nodes of a supersector are significant for impact, but centrality is not. In contrast, centrality is significant for response, suggesting that more central sectors are more likely to be affected by shocks. This is in line with what reported in [9], where it was observed that portfolios with stocks belonging to the peripheral region of information filtering networks are less risky. However, here we quantify this risk in terms of propagation of stress across the market, providing a tool to hedge risk and identify vulnerabilities within a reverse stress-testing framework.

Finally, we show how the information filtering network can be used to reverse stress test the system. More specifically, we identify the group of 10 nodes that collectively have the largest impact on the system. We find that these nodes correspond to funds, suggesting that stocks with a high level of underlying diversification have a higher propensity to impact the rest of the market.

By using the sparse probabilistic model of the whole market, we extract six temporal clusters that represent periods with different market behavior. The different clusters are identified from the similarity/dissimilarity in the likelihoods of the daily set of returns across all stocks (see ICC algorithm in [14]). We observe that there are important differences in the different market states, with periods of crisis showing a convergence of group behavior to the single node trend, indicating, in line with [15], convergent behavior in times of crisis.

We consider our approach to be a form of 'reverse stress testing' in a macroprudential sense. A strict microprudential definition of 'reverse stress testing' describes it as an exercise that involves exploring the size and nature of shocks that would render a bank's business model unviable, or its financial position fragile. It starts from an outcome of business failure and identifies circumstances where this might occur [16]. For the interested reader we point out that a summary of state-of-the-art approaches on MaPST, and roadmap for future research, was presented by the IMF in [17].

This approach to stress testing in financial systems is novel and presented for the first time in this paper. However, we make use of a number of tools and techniques developed mostly by some of the authors to retrieve, from observational data, the dependency structure in complex systems such as financial markets. Specifically, we build upon a data-driven modeling framework that estimates multivariate probabilities using the vast elliptical distribution family. We profit from information filtering tools that allow to sparsify such elliptical models over a meaningful network structure. We also take advantage of a methodology to construct temporal clusters of similar states in multivariate systems. All fundamental elements of these tools and techniques are sketched in the next section where all relative references are also reported.

2. Background

This paper builds upon several methodologies that have been developed in recent years, concerning four main areas:

(a) Modeling markets in terms of multivariate probability distributions;
(b) Use of networks to describe the interrelations between the variables in markets;
(c) Sparsification of the probabilistic models using these networks;
(d) Identification of different states of the market associated with different sparse multivariate probability distributions.

Hereafter, we briefly summarize the necessary state-of-the-art background in these area. For each of the topics, the reader interested in deepening into the subject is encouraged to refer to the original papers.

2.1. Modeling markets with multivariate elliptical distributions

The elliptical distribution family is a broad family of multivariate probability density functions that, notably, includes the multivariate normal and the multivariate student-t distributions. Multivariate elliptical

distributions are commonly used in modeling financial log-returns data [18–22] since they allow for the presence of heavy tails while inheriting many of the useful properties of the multivariate normal model [23]. In this paper, we indeed use the multivariate elliptical probability distribution to model the collective probabilistic structure of the log-returns for the entire market. We shall see shortly that, for the measure we have chosen as quantification of stress impact, there is no need to specify the kind of distribution within this broad family. Indeed, we argue that, within this multivariate probabilistic framework, stress is naturally modeled in terms of the multivariate conditional probability. Specifically, the probability distribution of the stressed variables is conditioned on the values of the stressing variables [13]. We show that the main contribution to stress is provided by the shift of the centroids of the stressed variables consequent to the applied stress. Such a shift depends on the covariance. The challenge is therefore to estimate accurately the inference structure and the parameter values of such multivariate probability distribution.

2.2. Networks in financial modeling

Information filtering networks are particularly effective for correlation-based graphs, and it has been shown that they describe well the structure of the market and its evolution with time [24, 25]. They have been used to characterise the spread of risk across a market [9], and it has been shown that risk does not distribute uniformly and the central or peripheral position of an asset in the market is an important risk factor.

It is common in the literature to find studies which link network centrality and systemic risk [26–29], and this has been exploited by policy makers in order to inform systemic risk rankings [30]. A direct focus on centrality measures in detecting systemically important financial institutions was applied by Kuzubaş *et al* [31], confirming ex-post that centrality values perform well in detecting systemically important institutions in interbank markets.

More general aspects of network structure have also been considered. For instance, Billo *et al* [32] show that indirect measures of firm interconnectedness based on principal component analyses and granger-causality networks are able to indicate periods of market dislocation and distress. In a similar vein, Diebold and Yilmaz [33] propose connectedness measures at all levels from system-wide to pairwise, and link this to marginal expected shortfall and conditional value at risk to emphasise the usefulness of these measures in a risk measurement and management setting.

2.3. Sparse modeling with information filtering networks

In this paper we make use of networks to represent the inference structure and better estimate the multivariate probability. Specifically, following the approach proposed in [34, 35], we construct sparse probabilistic models that have the information filtering networks encoded in the structure of the inverse covariance (i.e. partial correlations) [34]. Such a method is based on a special family of chordal networks which are clique-forests [36]. In particular, we use the triangulated maximally filtered graph (TMFG) [37], which is one instance of such a family. It is a three-clique structure made of tetrahedra separated by triangles, and it has the advantage to be computationally very efficient. It has been proven to be effective in identifying relevant data structures in different contexts from finance to psychology [38–40]. These networks can be used to estimate the maximal likelihood solution of elliptical multivariate distributions with sparse inverse covariance. This is a special solution for L_0-norm regularization, and it is a valid alternative of the popular Glasso method [41] for the covariance selection problem. Sparsification results in a better estimate of the covariance and overcomes issues related to the curse of dimensionality [35].

2.4. Identification of different market states

In this paper we use a recently developed time-clustering methodology (inverse covariance clustering, ICC, [14]) that combines the information filtering network description with a complete probabilistic modeling of the system to identify temporal clusters well represented by the same multivariate probability distribution. This method was chosen over other methods, as is efficient and reliable in identifying and predicting accurate and interpretable structures in multivariate, non-stationary financial datasets. This is in contrast to time series models such as TAR [42], which are often unable to identify structural breaks, and time series clustering techniques [43–45], which are highly susceptible to the curse of dimensionality. In [46], which considers equities traded in the US market, the method is able to distinguish a market state associated with both the 2008 crisis period and the COVID-19 as a distinct state from the long 'bull' period post 2008. ICC was also applied in a recent note in relation to the COVID-19 pandemic [46] to identify inherent market structures.

In this work, we similarly identify distinct states for the 2008 period, but for equities making up the FTSE 100 and 250. Pharasi *et al* [47] identify market states as clusters of similar correlation matrices applied to stocks making up the S & P 500 and Nikkei 225 indices. Similar correlation-based methods were applied by Münnix *et al* [48], who identify points of drastic change in correlation structure, which map to occurrences of financial crises. An alternative approach has been presented by Hendricks *et al* [49], in which a maximum likelihood

approach is applied to a physical analogy of the ferromagnetic Potts model at thermal equilibrium to cluster temporal periods as objects based on market microstructure feature performance.

3. Methodology

3.1. Conditional probability measure of systemic risk

From a probabilistic perspective the quantification of stress contagion between two sets of assets is measured by the conditional probability distribution, $P(\mathbf{Y}|\mathbf{X}=\mathbf{x})$, of the set of the stressed variables, $\mathbf{Y}$, under the condition that the set of stressing variables $\mathbf{X}$ is constrained at a given value of stress $\mathbf{x}$. Generally speaking, the conditioning can change both the kind of distribution and its parameters. For the vast elliptical distribution family, conditioning at $\mathbf{X}=\mathbf{x}$ causes a shift in the expected value of the conditioned variable. Here, we quantify stress contagion in terms of average loss in a set of assets caused by a loss imposed on another set of assets.

Consider the losses in the two sets of assets as represented by two multivariate sets of variables $\mathbf{X}\in\mathbb{R}^{p_\mathbf{X}\times 1}$ and $\mathbf{Y}\in\mathbb{R}^{p_\mathbf{Y}\times 1}$. Assuming they belong to the multivariate elliptical family probability distribution, then the conditional expected values are

$$\mathbb{E}[\mathbf{Y}|\mathbf{X}=\mathbf{x}]=\boldsymbol{\mu}_\mathbf{Y}+\boldsymbol{\Omega}_{\mathbf{YX}}\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{X}). \tag{1}$$

Where $\boldsymbol{\mu}_\mathbf{X}$ and $\boldsymbol{\mu}_\mathbf{Y}$ are the vectors of expected values of the variables $\mathbf{X}$ and $\mathbf{Y}$ respectively. The terms $\boldsymbol{\Omega}_{\mathbf{XX}}$, $\boldsymbol{\Omega}_{\mathbf{YX}}$ are the block elements of the shape matrix $\boldsymbol{\Omega}$

$$\boldsymbol{\Omega}=\begin{pmatrix}\boldsymbol{\Omega}_{\mathbf{XX}} & \boldsymbol{\Omega}_{\mathbf{XY}}\\ \boldsymbol{\Omega}_{\mathbf{YX}} & \boldsymbol{\Omega}_{\mathbf{YY}}\end{pmatrix}, \tag{2}$$

where $\boldsymbol{\Omega}_{\mathbf{XX}}$ is assumed invertible. In an applied scenario, the variables $\mathbf{X}$ and $\mathbf{Y}$ are log-returns of two sets of assets and equation (1) gives the expected log-returns of assets $\mathbf{Y}$ when variables $\mathbf{X}$ are stressed with values that deviate from the mean by $(\mathbf{x}-\boldsymbol{\mu}_\mathbf{X})$.

From equation (1), we observe that the effect of conditioning is shifting the centroids of the variables $\mathbf{Y}$ by $\boldsymbol{\Omega}_{\mathbf{YX}}\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}(\mathbf{x}-\boldsymbol{\mu}_\mathbf{X})$. It has been argued in [13] that such a shift is a good measure of systemic risk quantifying the average losses on variables $\mathbf{Y}$ when variables $\mathbf{x}$ deviate from the mean. Given the linearity in $\mathbf{x}-\boldsymbol{\mu}_\mathbf{X}$ of equation (1), the effect on the conditional expected values of $\mathbf{Y}$ is directly proportional to the size of the applied stress. We therefore propose to use the mean loss in variables $\mathbf{Y}$ when a stress equal to a unitary fluctuation from the mean $\mathbf{x}-\boldsymbol{\mu}_\mathbf{X}=\mathbf{1}_\mathbf{X}$ is applied. This is:

$$L_{\mathbf{X}\to\mathbf{Y}}=\frac{1}{p_\mathbf{Y}}\mathbf{1}_\mathbf{Y}^\top\boldsymbol{\Omega}_{\mathbf{YX}}\boldsymbol{\Omega}_{\mathbf{XX}}^{-1}\mathbf{1}_\mathbf{X}. \tag{3}$$

This measure of relative loss requires only the estimation of the shape matrix $\boldsymbol{\Omega}$. It therefore depends only on the structure of correlations between the variables and does not depend on the means. We shall refer to this quantity as 'impact', meaning that it qualifies the effect on the rest of the system ($\mathbf{Y}$) of a unitary stress applied on the group of variables $\mathbf{X}$. Similarly, we shall refer to $\boldsymbol{L}_{\mathbf{Y}\to\mathbf{X}}$ as 'response', and it quantifies the effect on $\mathbf{X}$ of a unitary stress applied on the rest of the system ($\mathbf{Y}$). In practical terms, equation (3) quantifies the propagation of average stress, indicating that a 100% stress from variables $\mathbf{X}$ will provoke on average $\boldsymbol{L}_{\mathbf{X}\to\mathbf{Y}}$% effect on variables $\mathbf{Y}$. Let us note that despite equation (1) quantifying a linear effect, does not assume linearity of the model. Indeed, it is valid for the whole elliptical distribution family of probability distributions, including the multivariate student-t which is 'fat-tailed' having therefore non-linear power law probability of large fluctuations away from the mean. Let us further notice that, in most practical cases, the probability of fluctuations away from the mean tends to decrease with conditioning and the principal contribution to risk is this shift of the mean under conditioning [13].

3.2. Identification of most impactful elements

Using the above measure, we investigated the group of most impactful variables by devising a simple algorithm that starts from a random group of n variables and then iteratively tries to replace one variable in the group with an external variable not in the group if such a replacement increases the total impact of the group. The procedure ends when it converges to a stable cluster. This simple procedure is not deterministic, and there are instances when the final result might change. We however verified that, in practice, in almost all cases, the same group of variables is selected. This simple procedure is not optimized for numerical efficiency, but for the purpose of this paper, where only a few hundreds of variables are involved, the algorithm converges in fraction of seconds on a standard laptop.

3.3. Construction of sparse inverse scale matrix using LoGo

We use the TMFG information filtering network [50] to estimate the L_0-norm regularized sparse inverse shape matrix $\mathbf{\Omega}$. Within the elliptical family of probability modeling, the shape matrix is proportional to the covariance (here assumed to be defined), and its sparse inverse is consequently proportional to the partial correlation. The zero entries in the sparse inverse shape matrix are therefore associated with zero conditional correlations. Note that however, beside normal models, such a zero conditional correlation does not imply conditionally independent variables. The L_0-norm regularized maximum likelihood shape matrix for the whole system is estimated from the local estimates of the shape matrices associated with the tetrahedral cliques in the TMFG, following the approach, called LoGo, outlined in [34]. This provides accurate and robust estimates that overcome the issue of the curse of dimensionality. Furthermore, it has been observed in [34, 51], that the off-sample likelihoods of the LoGo estimate are larger than the one obtained with the full-matrix max-likelihood estimation or other sparsification methods such as Glasso [41].

The LoGo sparse inverse covariance matrix is the network through which stress propagates. We identify the adjacency matrix of the market network structure with the non-zero elements in the LoGo sparse inverse covariance matrix. Later in the paper we will associate the centrality measure in such a network with the impact and response of stocks, investigating therefore the link between the dependency structure and the relative risk position.

3.4. Identification and clustering of market states

Financial market are not stationary systems. Their properties change with time and their log-returns in different time periods are described with different multivariate probabilities distributions. In order to capture these different market states and test our methodology across them, we clustered the time period by gathering together, in the same cluster, days which are described by similar multivariate probability distributions. The probabilistic description uses the multivariate elliptical distribution with a sparse inverse scale matrix. The clustering procedure follows [14], and is performed by starting from six samples composed of randomly gathered days. For each of these clusters, $c = 1, \ldots, 6$, we estimate the vector of means $\boldsymbol{\mu}_c$ and the sparse inverse covariance $\boldsymbol{J}_c = \mathbf{\Omega}_c^{-1}$. Then the following penalized log-likelihood is computed for each day:

$$\ell_{c,t} = \log|\boldsymbol{J}_c| - (\boldsymbol{X}_t - \boldsymbol{\mu}_c)\boldsymbol{J}_c(\boldsymbol{X}_t - \boldsymbol{\mu}_c)^T - \gamma\mathbf{1}(\mathcal{C}_{t-1} \neq c), \tag{4}$$

where $\mathbf{X}_t = (x_{t,1}, x_{t,2}, \ldots, x_{t,n})$ is the n-dimensional multivariate log-return vector for each day t; γ is a parameter penalizing state switching and; $\mathbf{1}(\mathcal{C}_{t-1} \neq c)$, is a penalizer function for cluster discontinuity that returns 1 if the cluster assignment of the observation at time $t-1$, $\mathcal{C}_{t-1}$, is different from the cluster assignment at time t. The clustering process is updated iteratively re-assigning days to clusters in a way to maximize $\sum_{t,c}\ell_{c,t}$. The procedure is made computationally efficient by using a Viterbi path. The process stops when a maximum is reached. This process is not deterministic, and it normally ends in a local maximum. However, the clustering structure in the different local maxima is usually rather similar. We re-run the whole procedure described in this paper for ten times and we observed that, across all ten runs, the final results and conclusions are consistent with those reported hereafter.

4. Data

We consider end of day prices for stocks making up the FTSE 100 and 250 indices, from January 2005 to August 2020. Of these stocks, 231 of the 350 were available across the whole period analyzed; those that were not, were not considered. This time period contained two periods of greater market stress—the 2008 financial crisis, and the initial market shocks experienced in response to the COVID-19 pandemic. Making use of the global industry classification standard, these stocks were classified into 11 sectors. The number of stocks per sector can be seen in table 1. Some stocks in the raw data were labeled with N/A as the sector; upon inspection it was found that these were all some form of financial fund, and so these stocks were relabeled as 'funds'.

5. Results

5.1. Market states

The clustering of the system into six market states associated with daily maximal likelihood yields the partition represented in figure 1. The choice of six has been mostly guided by the observation that it provides a good distinction between the various periods. The results have been obtained using $\gamma = 100$, which provides average sizes of uninterrupted clusters of about one month. One can observe that the six clusters spread unevenly through the observation window. The ordering follows their average position in time, with cluster 1 mostly

Table 1. Number of stocks per GICS SuperSector.

Sector	Number of stocks
Industrials	47
Funds	43
Consumer discretionary	29
Financials	25
Real estate	20
Materials	15
Consumer staples	14
Communication services	10
Information technology	10
Utilities	7
Health care	5
Energy	4

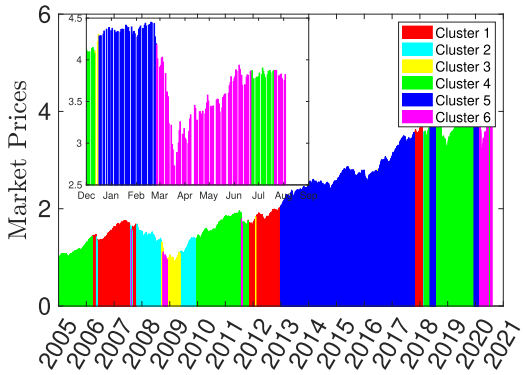

Figure 1. Market states obtained through the clustering algorithm described in section 3.4. The *y*-axis reports the average price across the market (daily data) and the colors represent the six clusters.

present at the early stages between 2005 and 2013 while cluster 6 is mostly present during the 2020 Covid-related market turmoil. Note that we observe days in the 2007–2009 crisis period being also associated with cluster 6, however most of that crisis period is associated with cluster 3. As already mentioned, this clustering method is not deterministic and different runs can return different clustering. Therefore, for this study, the clustering was executed ten times and all results presented in this paper were computed for each instance.

5.2. Response and impact versus network centrality

We investigate the relation between impact and response of stocks with respect to their centrality within the information filtering network. At the level of single nodes we observe that more central nodes have both higher impact and response. This is demonstrated by the dark black lines in figures 2–4, which report the average impact and response of the stocks vs average centrality, displaying an increasing trend in both cases. However, when we look at groups of nodes we observe that the most central clusters have a higher response but lower impact. Here the group centrality is simply the mean centrality of the group constituents. This is demonstrated by the cyan lines in figures 2–4, which report the average impact and response of group of stocks with various sizes. Also the same plots for the group of stocks corresponding to industry sectors reveal a consistent trend. This effect is most likely a consequence of the fact that central clusters are compact and therefore have a large internal effect on each other—which is not accounted for in the impact measure, and are less impactful on the rest of the system.

This is illustrated in figure 2, where we report impact/response as function of the centrality of single nodes (black lines) and the average impact/response vs average centrality for random groups of nodes with different sizes (cyan lines). These data refer to the whole period. Specifically, in the top plot in figure 2, the lines show that centrality increases the response effect with little influence from the grouping. Conversely, the bottom plot of figure 2 demonstrates that single nodes (black lines) are also increasing their impact with centrality but, when grouped, the opposite happens and greater centrality corresponds to smaller impact (cyan lines). In these figures symbols report the average impact/response vs average centrality for the super-sectors. We see an overall consistency with the results for random groupings however with the sectors being overall a bit less responsive and impactful than the random groupings. The analyses over the market states reported in

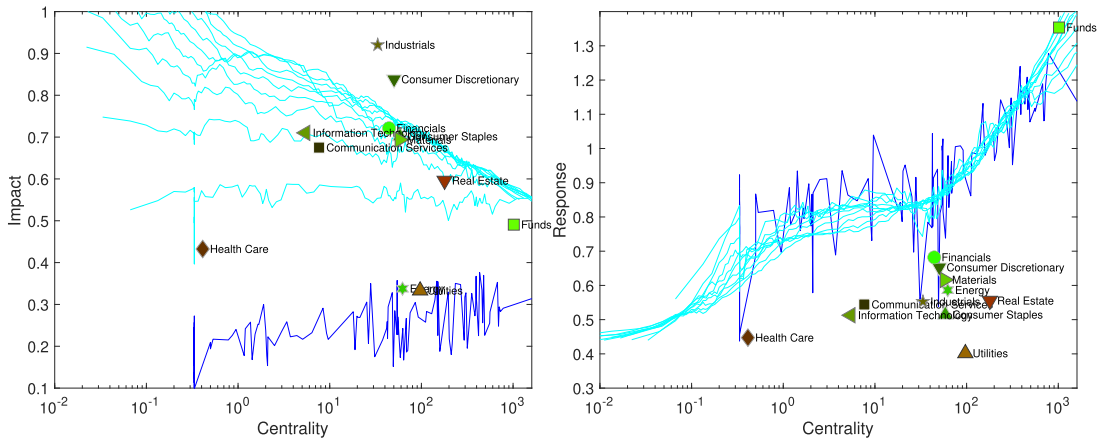

Figure 2. Impact/response vs centrality computed over the whole period. The darker black line correspond to single nodes, cyan lines are group of nodes form 5 to 50. Symbols represent the supersectors.

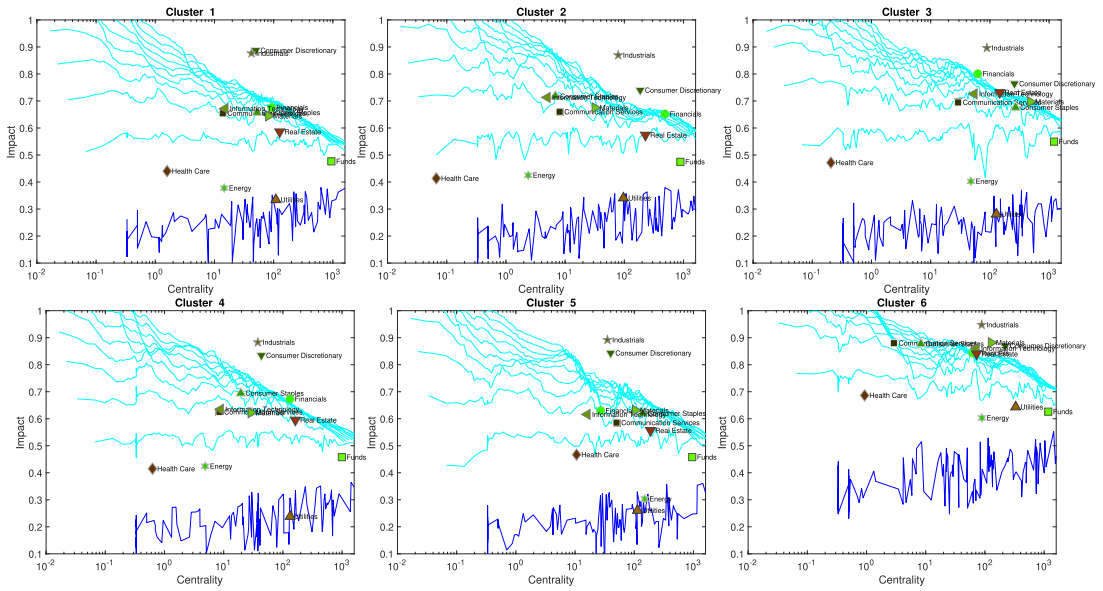

Figure 3. Impact vs centrality for the six clusters. The black line corresponds to single nodes, cyan lines are group of nodes form 5 to 50. Symbols are the supersectors.

figures 3 and 4 give similar overall results with some significant differences in different market periods. When considering the impact scores, we see a clear increase in impact in cluster 6, both in the single and grouped node trends. We also see this to some extent in cluster 3, but this is limited to only some sectors. A more evident change is seen in the response scores of the supersectors in the two periods of crisis, which are dominated by market states three and six. In both periods of crisis, the funds sector has a lower response, with the majority of the other sectors seeing an upward shift in response score. In all cases, the sector specific response scores are seen to move closer to the single node and grouped node trend. Some sectors are more affected in times of crises than others, for example the utilities and healthcare sectors show larger increases in response and centrality. We also see a large increase in centrality on average across sectors in cluster 5.

5.3. Response and impact of supersectors

The results of section 5.2 show that, while more central nodes have both higher response and impact, if we aggregate nodes into random groups, more central groups have higher response, but lower impact. Results are less clear when we group nodes into supersectors.

Here we take a closer look to the aggregation of firms into supersectors, and we perform a regression analysis to understand if the impact and response measures of supersectors are affected by their centrality.

There are two apparent differences between the random groups of sections 5.2 and supersectors: first, random groups, associated with each of the cyan lines have all the same size, while the size of supersectors ranges

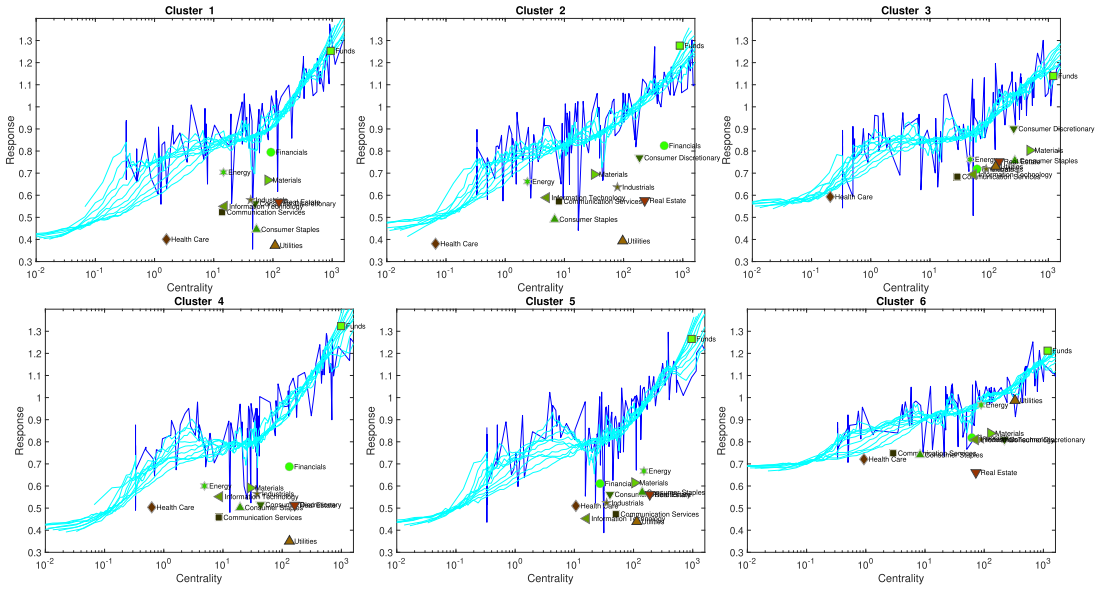

Figure 4. Impacted vs centrality for the six clusters. Black line corresponds to single nodes, cyan lines are group of nodes form 5 to 50. Symbols are the supersectors.

Table 2. Linear regression for the measure of impact ($R^2 = 0.53$).

Variable	Value	p-value
Size	1.33	1.2×10^{-12}
Fraction of links within supersector	−1.06	7×10^{-8}
Log centrality	−0.02	0.86

Table 3. Linear regression for the response measure ($R^2 = 0.52$).

Variable	Value	p-value
Size	−0.34	0.03
Fraction of links within supersector	0.79	3×10^{-5}
Log centrality	0.25	0.026

from 4 to 47. Secondly, nodes in random groups tend to be dispersed in the network, while nodes within the same supersector tend to be closer to each other. In order to account for this, we run a regression where, in addition to considering centrality as an independent variable, we control for the size of the supersector and the fraction of a supersector's links that are shared by nodes within the supersector.

The results are reported for one realization of the network in tables 2 and 3 for impact and response measures respectively.

We see from table 2 that the coefficients associated with size and fraction of links shared by nodes of a supersector are statistically significant. The positive sign of the coefficient corresponding to the size is simply due to the fact that as the size of the supersector increases the shock affects a higher number of nodes. The negative sign of the fraction of links connecting nodes within the supersector is instead due to the fact that links between nodes that are exogenously shocked do not increase the impact, so that the more compact a supersector is, the less it affects the others. We also observe that the centrality of supersectors does not appear to play a significant role here, and that the effect of the network structure is simply accounted for by the fraction of links within supersectors.

This is not the case for the response measure. We see from table 3 that centrality remains statistically significant (p-value 2.6%, see table 3) with a positive coefficient, signaling the tendency of more central sectors to show a higher response. We also observe a change in the sign of the coefficients associated with the other two variables. In particular, we see that the fraction of links shared by nodes of a supersector now has a positive coefficient, denoting the fact that more compact supersectors tend to have a higher response. This is due to the

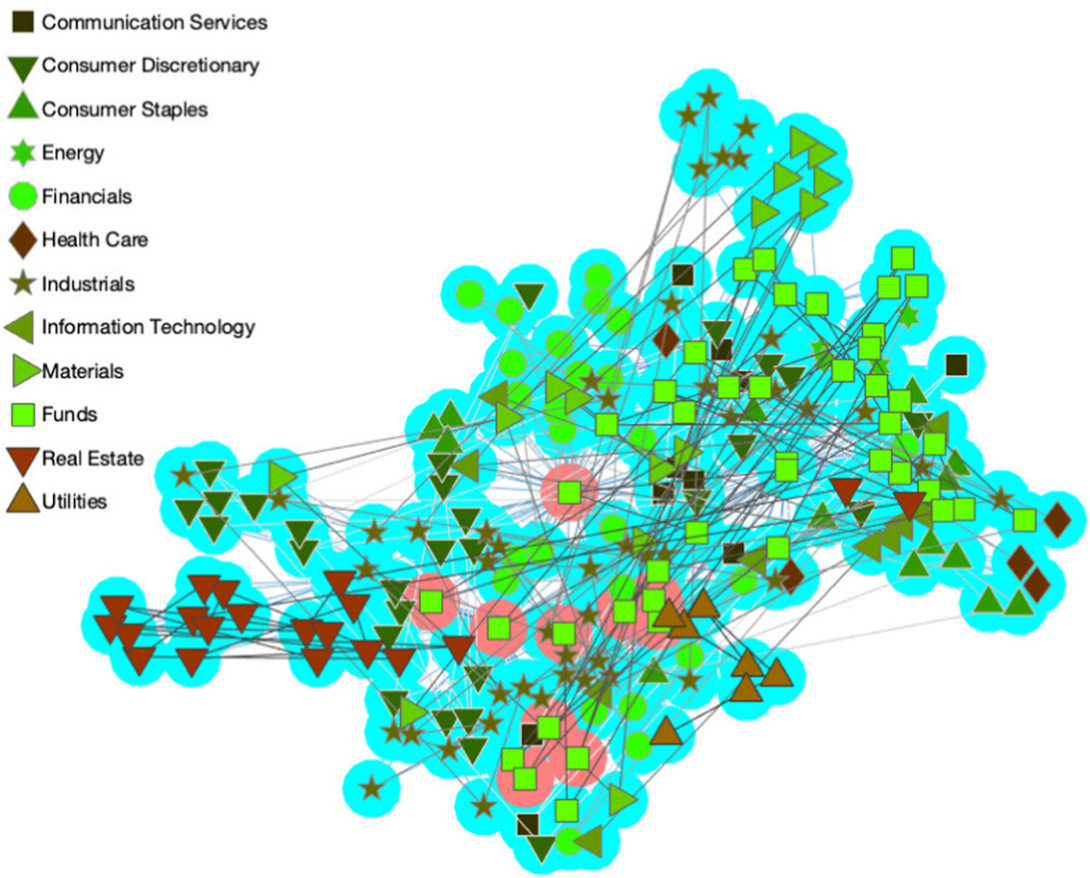

Figure 5. Representation of the market network corresponding to non-zero elements of the inverse covariance computed over the whole-period. The vertices correspond to individual stocks. Their industry sector is indicated with different symbols accordingly with the legend and the group of 10 stocks with the collective highest impact is highlighted in with the red disks.

Table 4. Groups of 10 stocks with highest impact across the six different time-clusters. *Capital goods industrials, **diversified financials, all others funds.

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
ATS	CI	FCI	FGT	SDR**	ICP**
MRC	ATS	ATS	MRC	FGT	MRC
WTA	TEM	MRC	CTY	MRC	CTY
CTY	CTY	WTA	TRY	WTA	TRY
TRY	MYI	CTY	EDI	CTY	EDI
EDI	JMG	FEV	ASL	TRY	ASL
ASL	FEV	CLD	HSL	EDI	BRS
AGT	EDI	ASL	TMP	ASL	HSL
TMP	TMP	IMI*	PLI	TMP	TMP
PLI	PLI	LWD**	FSV	PLI	PLI

fact that, at odds with the case of impact, the contribution of these links to loss propagation is accounted for in the calculation of the response variable.

5.4. Nodes with the highest impact

Concerning the identification of the nodes with the highest impact, we find that the most influential $n = 10$ nodes for the whole period are: MRC, WTA, CTY, EDI, ASL, BRS, HSL, TMP, PLI, FSV, which all belong to the funds supersector. Their position in the network is highlighted in figure 5. We see that these stocks do not localise to a particular region of the network, suggesting that they are not necessarily displaying similar behaviors leading to this higher level of impact. Note that this is the group of 10 stocks with collective highest impact, not the collection of single highest impacting stocks.

We observe similar results across the different market states, as summarized in table 4. Firstly, we note again that the majority these stocks belong to the funds supersector, suggesting that stocks with a high level of diversification are likely to impact the rest of the market the most when experiencing shocks. We also notice that

there are several repetitions in the groups of highest impact stocks across the six different periods. For instance, CTY appears in all six clusters and there are five other stocks (MRC, EDI, ASL, TMP, PLI) that appear in five out of six clusters. They also all appear in the full-period group reported in figure 5. Cluster 3 is the one with least overlap with the rest of the clusters indicating that the risk scenario during the 2007–2009 crisis was quite special. Conversely, cluster 6 has similar overlaps as the other clusters suggesting that the 2020 Covid-related crisis is instead less peculiar.

6. Conclusion

In this paper, we have presented a novel method to measure the impact and response of stocks in a market when shocks are experienced. For the first time, we associate the structure of the information filtering network with a quantitative risk measure relating the behavior of impact and response to the structure of the underlying inverse covariance matrix of the stock log-returns. In the application to different market states observed from 2005 to 2020 in FTSE 100 and 250 markets, we observe that both impact and response are related to the centrality of individual nodes in the network, but that central groups have a higher response and lower impact due to the internal effects for these groups.

We observe markedly different behaviors in different states of the market, particularly in the 2008 and COVID-19 crises. The convergence to the single and grouped node trend observed for response scores during periods of crisis is consistent with observations that correlations between all stock prices are seen to increase in these periods. This is not only consistent with our own observations but also with recent observations made by Sandoval *et al* [15]. This tendency of 'markets to behave as one' in times of crisis is interesting to monitor from the context of response and impact scores—particularly in the case of sectors which, in a period of comparative market stability, are seen to have a lower value according to response than the individual/group trend. These sectors might be expected to respond worse to the crisis period, whereas the effect on sectors closer to the trend, or above, may see a lesser impact. Moreover, we observe the financial funds sector to show the largest response to shocks across all periods, which suggests that the higher levels of diversification of these stocks makes them more susceptible to shocks. On the other hand, the industrials sector shows the highest impact, however this is likely due to this being the largest of the sectors. Some sectors are more affected in times of crises than others, for example the utilities and healthcare sectors show larger increases in response and centrality.

Since it relies on historical returns, our methodology is only able to compute impact and response of firms provided that the future structure of dependencies among variables has been observed in the past. This limitation is mitigated by the fact that we consider different time clusters, allowing us to identify different market states, which can then be used to perform stress tests under different scenarios for what concerns the structure of dependencies between stocks.

The analysis of time-clusters also provides a test for the robustness of the main result concerning the link between network centrality and impact (or response). Indeed, despite specific differences in the six time-clusters, the overall main finding, that individual stock impact and response increase with centrality but group impact decreases with centrality, reproduces similarly in all clusters and in the whole period. Note that these are samples of different sizes and with data from different periods.

Our work can be extended in several directions. For instance, the robustness of our findings for what concerns the relationship between centrality, impact and vulnerability could be further tested on different markets. Furthermore, we measured here the impact of variable X in terms of the average effect that stressing this variable will cause on variable Y. While the main contribution to the propagation of stress is in fact due to the shift of the conditional distribution of Y, our methodology would also allow to compute more complex quantities such as, for instance, average losses above a given quantile. Finally, it would be interesting to compare results from historical periods of market turmoil associated with different types of threats (e.g. disease outbreak vs economic vs geopolitical vs technological catastrophes [52]). Finally, we point out that the methods discussed in this paper can be applied to contexts other than the study of financial markets, such as e.g. the propagation of reputational risk between publicly traded firms, or the response of ecosystems to environmental disturbances, and in general to any system where dependencies between variables can be inferred from multivariate time series.

Acknowledgments

The authors acknowledge discussions with many members of the Financial Computing and Analytics group at UCL. In particular a special thank to Guido Massara, Carolyn Phelan and Pier Francesco Procacci. Also, thanks for support from ESRC (ES/K002309/1), EPSRC (EP/P031730/1) and EC (H2020-ICT-2018-2 825215).

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Tomaso Aste https://orcid.org/0000-0002-4219-0215

References

[1] Caccioli F, Barucca P and Kobayashi T 2018 Network models of financial systemic risk: a review *J. Comput. Soc. Sci.* **1** 81–114
[2] Bardoscia M, Barucca P, Battiston S, Caccioli F, Cimini G, Garlaschelli D, Saracco F, Squartini T and Caldarelli G 2021 The physics of financial networks *Nat. Rev. Phys.* **3** 490–507
[3] Albert R, Albert I and Nakarado G L 2004 Structural vulnerability of the north American power grid *Phys. Rev.* E **69** 025103
[4] Pagani G A and Aiello M 2013 The power grid as a complex network: a survey *Physica* A **392** 2688–700
[5] Albert R, Jeong H and Barabási A-L 2000 Error and attack tolerance of complex networks *Nature* **406** 378–82
[6] Cohen R, Erez K, Ben-Avraham D and Havlin S 2001 Breakdown of the internet under intentional attack *Phys. Rev. Lett.* **86** 3682
[7] Dunne J A 2006 The network structure of food webs *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford: Oxford University Press) pp 27–86
[8] Allesina S and Pascual M 2009 Googling food webs: can an eigenvector measure species' importance for coextinctions? *PLoS Comput. Biol.* **5** e1000494
[9] Pozzi F, Di Matteo T and Aste T 2013 Spread of risk across financial markets: better to invest in the peripheries *Sci. Rep.* **3** 1665
[10] Song W-M and Zhang B 2015 Multiscale embedded gene co-expression network analysis *PLoS Comput. Biol.* **11** e1004574
[11] Kenett Y N, Anaki D and Faust M 2014 Investigating the structure of semantic networks in low and high creative persons *Front. Hum. Neurosci.* **8** 407
[12] Alexander-Bloch A F, Gogtay N, Meunier D, Birn R, Clasen L, Lalonde F, Lenroot R, Giedd J and Bullmore E T 2010 Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia *Front. Syst. Neurosci.* **4** 147
[13] Aste T 2020 Stress testing and systemic risk measures using multivariate conditional probability (arXiv:2004.06420)
[14] Procacci P F and Aste T 2019 Forecasting market states *Quant. Finance* **19** 1491–8
[15] Sandoval L and Franca I 2011 Correlation of financial markets in times of crisis *Physica* A **391** 187–208
[16] Bailey A *et al* 2020 *Financial Stability Report* (Bank of England Financial Policy Committee) p 45
[17] Anderson R, Danielsson J, Baba C, Das U S, Kang H and Segoviano M 2018 Macroprudential stress tests and policies: searching for robust and implementable frameworks *IMF Working Paper* International Monetary Fund p WP/18/197
[18] Flood M D and Korenko G G 2015 Systematic scenario selection: stress testing and the nature of uncertainty *Quant. Finance* **15** 43–59
[19] Owen J and Rabinovitch R 1983 On the class of elliptical distributions and their applications to the theory of portfolio choice *J. Finance* **38** 745–52
[20] Kring S, Rachev S T, Höchstötter M, Fabozzi F J and Bianchi M L 2009 Multi-tail generalized elliptical distributions for asset returns *Econom. J.* **12** 272–91
[21] Christodoulakis G A and Satchell S E 2006 Exact elliptical distributions for models of conditionally random financial volatility *Working Papers 32* (Bank of Greece)
[22] Yang X, Zheng X and Chen J 2021 Testing high-dimensional covariance matrices under the elliptical distribution and beyond *J. Econom.* **221** 409–23
[23] Frahm G 2004 Generalized elliptical distributions: theory and applications *PhD Thesis* Universität zu Köln
[24] Tumminello M, Aste T, Di Matteo T and Mantegna R N 2005 A tool for filtering information in complex systems *Proc. Natl Acad. Sci. USA* **102** 10421–6
[25] Aste T, Shaw W and Di Matteo T 2010 Correlation structure and dynamics in volatile markets *New J. Phys.* **12** 085009
[26] Markose S, Giansante S and Shaghaghi A R 2012 'Too interconnected to fail' financial network of US CDS market: topological fragility and systemic risk *J. Econ. Behav. Organ.* **83** 627–46
[27] Battiston S, Puliga M, Kaushik R, Tasca P and Caldarelli G 2012 Debtrank: too central to fail? Financial networks, the fed and systemic risk *Sci. Rep.* **2** 541
[28] Martinez-Jaramillo S, Alexandrova-Kabadjova B, Bravo-Benitez B and Solórzano-Margain J P 2014 An empirical study of the Mexican banking system's network and its implications for systemic risk *J. Econ. Dyn. Control* **40** 242–65
[29] Bravo-Benitez B, Alexandrova-Kabadjova B and Martinez-Jaramillo S 2016 Centrality measurement of the Mexican large value payments system from the perspective of multiplex networks *Comput. Econ.* **47** 19–47
[30] De Bruyckere V 2015 Systemic risk rankings and network cnetrality in the European banking sector *ECB Working Paper* European Central Bank
[31] Kuzubaş T U, Ömercikoğlu B and Saltoğlu I 2014 Network centrality measures and systemic risk: an application to the Turkish financial crisis *Physica* A **405** 203–15
[32] Billio M, Getmansky M, Lo A W and Pelizzon L 2012 Econometric measures of connectedness and systemic risk in the finance and insurance sectors *J. Financ. Econ.* **104** 535–59
[33] Diebold F X and Yilmaz K 2014 On the network topology of variance decompositions: measuring the connectedness of financial firms *J. Econom.* **182** 119–34
[34] Barfuss W, Massara G P, Di Matteo T and Aste T 2016 Parsimonious modeling with information filtering networks *Phys. Rev.* E **94** 062306
[35] Aste T 2020 Topological regularization with information filtering networks (arXiv:2005.04692)
[36] Massara G P and Aste T 2019 Learning clique forests (arXiv:1905.02266)
[37] Massara G P, Di Matteo T and Aste T 2016 Network filtering for big data: triangulated maximally filtered graph *J. Complex Netw.* **5** 161–78
[38] Nicola G, Cerchiello P and Aste T 2020 Information network modeling for US banking systemic risk *Entropy* **22** 1331

[39] Turiel J D, Barucca P and Aste T 2020 Simplicial persistence of financial markets: filtering, generative processes and portfolio risk (arXiv:2009.08794)
[40] Christensen A P, Kenett Y N, Aste T, Silvia P J and Kwapil T R 2018 Network structure of the Wisconsin Schizotypy scales-short forms: examining psychometric network filtering approaches *Behav. Res.* **50** 2531–50
[41] Friedman J, Hastie T and Tibshirani R 2008 Sparse inverse covariance estimation with the graphical lasso *Biostatistics* **9** 432–41
[42] Tong H 1978 On a threshold model *Pattern Recognition and Signal Processing* (Dordrecht: Springer)
[43] Ren L, Wei Y, Cui J and Du Y 2017 A sliding window-based multi-stage clustering and probabilistic forecasting approach for large multivariate time series data *J. Stat. Comput. Simul.* **87** 2494–508
[44] Nevill-Manning C G and Witten I H 1997 Identifying hierarchical structure in sequences: a linear-time algorithm *J. Artif. Intell. Res.* **7** 67–82
[45] Liao W T 2005 Clustering of time series data—a survey *Pattern Recognit.* **38** 1857–74
[46] Procacci P, Phelan C and Aste T 2020 Market structure dynamics during Covid-19 outbreak (arXiv:2003.10922v1 [q-fin.ST])
[47] Pharasi H K, Seligman E and Seligman T H 2020 Market states: a new understanding (arXiv:2003.07058)
[48] Münnix M, Shimada T, Schäfer R, Leyvraz F, Seligman T H, Guhr T and Stanley H 2012 Identifying states of a financial market *Sci. Rep.* **2** 644
[49] Hendricks D, Gebbie T and Wilcox D 2015 Detecting intraday financial market states using temporal clustering *Quant. Finance* **16** 1657–78
[50] Massara G P, Di Matteo T and Aste T 2017 Network filtering for big data: triangulated maximally filtered graph *J. Complex Netw.* **5** 161
[51] Procacci P F and Aste T 2021 Portfolio optimization with sparse multivariate modelling (arXiv:2103.15232)
[52] Coburn A, Ralph D, Tuveson M, Ruffle S and Bowman G 2013 A taxonomy of threats for macro-catastrophe risk management *Working Paper* (Cambridge: Centre for Risk Studies, University of Cambridge) pp 20–4