Network Structure and Market Risk in the European Equity Market

Germán G. Creamer, School of Business, Stevens Institute of Technology

ABSTRACT

This paper proposes a methodology to anticipate market risk using qualitative and quantitative variables that capture communicative and financial activity within equity networks. During periods of crisis as market risk increases, companies tend to behave alike, and the number of news and common topics among companies increases. I built a corporate news network where the nodes are top European companies, and the edges are the number of news items on the same topic by every pair of companies identified by the topic model methodology. I conducted a longitudinal analysis using a time series of static social networks to generate a dynamic social network and proposed the component causality index as a leading indicator of market risk. This research finds out that the component causality index, based on centrality indicators, anticipates or moves together with Value-at-Risk (VaR) during the period 2005-11.

KEY WORDS

Social networks, text analysis, link mining, risk management, market risk, systemic risk

I. Introduction

After the final crisis of 2008, the new risk literature has explored the linkages among financial institutions and corporations as drivers of the crisis as well as indicators of financial stress. An important part of these models is based on overnight payments among financial institutions or correlations among returns of different institutions [1, 2, 3, 4]. The main idea is that these measures may overcome the restrictive perspective of the standard measures of market risk endorsed by the Basel accords such as Value at Risk (VaR) or Conditional VAR (CVaR) which are only based on financial profits (Section II).

Another recent line of research is forecasting using economic networks such as the product network proposed by Dhar et al. [5] to forecast product demand. In finance, Creamer and Stolfo [6] applied a link mining algorithm called CorpInterlock to integrate the metrics of an extended corporate interlock (social network of directors and financial analysts) with corporate fundamental variables and analysts' predictions (consensus) to forecast the trend of the cumulative abnormal return and earnings surprise of US companies. Creamer et al. [7] also predicted return and volatility using a corporate network of European companies based on the common topics of corporate news, and Adamic et al. [8] used a traders network to quantify the flow of information through financial markets. Some recent papers [9, 10, 11], explored the effect of network variables on market and systemic risk.

This paper integrates both perspectives. Diverse corporate networks and companies structure may have an effect on asset price dissemination, volatility, and risk. However, modern finance theory does not take into account the network structure and dynamics that define assets return and volatility. I propose a method to anticipate the market risk in the equity market using time series, text analytics and network indicators. This systemic approach evaluates the impact of several centrality indicators of a corporate news network on market risk, and test it using the European STOXX 50 index. The proposed method includes a unique text analytics component based on common topics from the news that can help to anticipate future instability of the financial market. Therefore, this method can complement other measures of financial stress or systemic risk such as those described by [9, 12, 13] that do not include this perspective.

II. Market Risk

For Basel III, the financial risk includes credit risk, market risk, and operational risk. The market risk formula for Basel III uses 10-day 99% VaR and stressed VaR (99.9%) [14]. VaR is the maximum loss of an asset or a portfolio with a given confidence level in a certain period. Conditional VaR (CoVaR) or expected shortfall has become an alternative and preferred measure of risk as is the conditional expected value of the loss larger than VaR for a certain horizon and confidence level. Both VaR and CoVaR are calculated using the nonparametric and the parametric approach. The nonparametric approach leads to the following calculations of VaR and CoVaR (see [15, 16, 17]):

$$VaR(\alpha) = -\hat{q}(\alpha)W$$

Manuscript received ...

The author would like to thank Diccon Close, SIRCA, and Thomson Reuters NewsScope for providing access to their news' archives, and Yong Ren for his research assistance. The opinions presented are the exclusive responsibility of the author. A preliminary version of this paper was presented at the Workshop on Information Networks (WIN) 2015 at New York University. This work was partially supported by the Stevens Alliance for Technology Management.

G. Creamer is with the School of Business, Stevens Institute of Technology, NJ. gcreamer@stevens.edu

$$CoVaR = \frac{\int_{i=1}^{n} L_i I\{L_i > VaR(\alpha)\}}{\int_{i=1}^{n} I\{L_i > VaR(\alpha)\}}$$

where $L_i = -R_i W$, $\hat{q}(\alpha)$ is the α-quantile of the series of historical returns $R_1, ..., R_n$, $(1-\alpha)$ is the confidence level, W is the amount invested in the asset under study, and $I\{L_i > VaR(\alpha)\}$ is the indicator function for $L_i > VaR(\alpha)$. The negative sign transforms revenue to a loss.

VaR and CoVaR under the parametric approach are the following:

$$VaR(\alpha) = -WF^{-1}(\alpha|\hat{\theta})$$

$$CoVaR = -\frac{W}{\alpha} \int_{-\infty}^{F^{-1}(\alpha|\hat{\theta})} x f(x|\hat{\theta}) dx$$

where $F(y|\theta)$ and $f(x|\hat{\theta})$ are the cumulative and probability distribution functions respectively of the return distribution calibrated with the parameters θ. This distribution is typically normal although it could follow other distributions such as the Student-t to simulate the fat tails that are characteristic of the volatility series.

Using a normal distribution, $VaR(\alpha) = (-\mu + z_{1-\alpha}\sigma)W$ where $z_{1-\alpha}$ is the z value of the standard normal distribution $Z \sim N(0,1)$ with a $(1-\alpha)$ confidence level, and σ is the standard deviation of the particular asset return [18]. In practical terms and assuming that daily returns are close to zero, daily VaR is calculated in the following way: $VaR(\alpha) = z_{1-\alpha}\sigma W$. The latter is the approach used in this paper to calculate VaR.

The need to perform these risk calculations to fulfill corporate legal requirements sets up an interesting research question with a large scale social impact: How should models, based on empirical financial and non-financial data, be designed that help institutions and individuals best understand, anticipate, and mitigate risks?

The current financial time series econometric models may not always capture extreme events generated by the complex and chaotic nature of the forces acting on financial markets. For example, during the past 120 years, there have been a steady stream of financial and banking crises, stock market bubble bursts, and credit crunches that seem to happen ever more often. The effects of these catastrophic events are intensified by the fact that the contemporary time series econometric models often fail to anticipate sudden changes in data, leaving participants exposed to huge losses. The ramifications of these losses have large economic and social impact. In response to these limitations, recent risk models based on social networks and complex adaptive systems have emerged and shown that social network indicators can be used for financial prediction [19, 20, 21, 22].

This paper answers the above question proposing a method to extract indicators from a corporate network based on news' common topics to anticipate major changes in the market volatility and risk. In particular, for the analysis of market risk, I propose an index called Component Causality Index (CCI) which is the proportion of the components of a particular system or index that have significant causal relationships with a dependent variable over a given period. In the case of this research, the components are the companies of the STOXX 50 index and the dependent variable is the 99.9% VaR of the STOXX 50 index.

The main idea is that if there are important changes in the components of a system or an index, the VaR or volatility of the system will also be affected, and therefore could be anticipated by the change of behavior of its components. I use the CCI as a leading indicator of market risk evaluating the impact of the network variables on the next period VaR for the complete time series.

III. Technical Approach

A. Methods

In this section, I describe the following methods used to build corporate news networks and evaluate the causality among the main time series under analysis.

1) Granger causality: Granger causality [23] is a very popular methodology used in economics, financial econometrics, as well as in many other areas of study, such as neuroscience, to evaluate the linear causal relationship between two or more variables. According to the basic definition of Granger causality, the forecasting of the variable Y_t with an autoregressive process using Y_{t-l} as its lag-l value should be compared with another autoregressive process using Y_{t-l} and the vector X_{t-l} of potential explanatory variables. Thus, X_{t-l} Granger causes Y_t when X_{t-l} happens before Y_t, and X_{t-l} has unique information to forecast Y_t that is not present in other variables.

Typically, Granger causality is tested using an autoregressive model with and without the vector X_{t-1}, such as in the following bivariate example:

$$Y_t = \sum_{l=1}^{L} \alpha_l Y_{t-l} + \epsilon_1 \quad (1)$$

$$Y_t = \sum_{l=1}^{L} \alpha_l Y_{t-l} + \sum_{l=1}^{L} \beta_l X_{t-l} + \epsilon_2 \quad (2)$$

where the residual ϵ_j is a white noise series: $\epsilon_j \sim N(0, \sigma)$, j=1,2.

X_{t-l} Granger causes Y_t if the null hypothesis $H_0 : \beta_l = 0$ is rejected based on the F-test. The order of the autoregressive model is selected according to either the Akaike information criterion or the Bayesian information criterion. This research applies a multivariate Granger causality test to evaluate the simultaneous effects of several variables in volatility [24].

2) Brownian distance and distance correlation test of independence: Székely and Rizzo [25] proposed a multivariate nonlinear dependence coefficient called Brownian distance correlation that can be used with random vectors of multiple dimensions or with strongly stationary time series. These authors also proposed the Brownian distance covariance, which captures the covariance on a stochastic process. Distance covariance between the random vectors X and Y measures the Euclidean distance between f_X, f_Y and $f_{X,Y}$ where f_X and f_Y are the characteristic functions of X and Y respectively, and $f_{X,Y}$ is the joint characteristic function of X and Y.

Székely and Rizzo [26] observed that the distance correlation in a high dimensional space goes to 1 even for a pair of independent variables. Therefore, they proposed a modified distance correlation which, under independence and in high dimensions, tends to converge to an approximately normal Student t distribution. The distance correlation (R^*) can take negative values and $|R^*| \leq 1$. The distance correlation test of independence is based on the following transformation of R^*:

$$\tau_n = \sqrt{v-1}\frac{R_n^*}{1-\sqrt{R_n^*}} \tag{3}$$

where $v = \frac{n(n-3)}{2}$ and n is the sample size. τ converges to a Student t distribution with $v-1$ degrees of freedom.

In this paper, I evaluate the non-linear dependence of any financial time series such as the current value of Y (Y_t) on the l lagged value of the matrix of $\mathbf{X}$ ($\mathbf{X_{t-1}}$) dependent variables with the distance correlation $R^*(\mathbf{X_{t-1}}, Y_t)$. In particular, I wish to explore the lead-lag relationship among the time series under study. If $R^*(\mathbf{X_{t-1}}, Y_t) \neq 0$ and $l > 0$, then $\mathbf{X_{t-1}}$ leads the series Y_t. Additionally, if $R^*(\mathbf{X_{t-1}}, Y_t) \neq 0$, $R^*(\mathbf{X_t}, Y_{t-l}) = 0$ and $l > 0$, then there is an unidirectional relationship from $\mathbf{X_{t-1}}$ to Y_t. However, if $R^*(\mathbf{X_{t-1}}, Y_t) \neq 0$, $R^*(\mathbf{X_t}, Y_{t-l}) \neq 0$ and $l > 0$, then there is a feedback relationship between $\mathbf{X}$ and Y. On the contrary, if $R^*(\mathbf{X_{t-1}}, Y_t) = 0$ and $R^*(\mathbf{X_t}, Y_{t-l}) = 0$ then there is no lead-lag relationship between $\mathbf{X}$ and Y [27]. In this research, I only evaluate the dependence of volatility on several other variables. However, as I calculate volatility using a GARCH(1,1) model then the autocorrelation of the squared residuals has already been removed as confirmed by the Ljung-Box test, (see Table I-b, 2nd. line).

The distance correlation for the multivariate case calculates the overall effect of all the variables on the dependent variable. Therefore, it is necessary to run stepwise tests where variables are added and removed to evaluate their contribution to the multivariate correlation between the independent and the dependent variables. This research partially follows this approach as it tests each variable and several meaningful sets of dependent variables. This approach also considers any indirect effects that may exist among these variables.

3) Topic model: The topic model methodology [28] discovers common topics among a series of documents. This approach assumes that documents are a mixture of topics, where a topic is based on a probability distribution over words. Topics are chosen according to their distribution, and keywords are associated with specific topics to evaluate a new document. By inverting this process, it is possible to infer the set of topics used to generate the documents. The Latent Dirichlet Allocation (LDA) is an accepted topic model methodology for capturing the latent structure of a large set of documents. LDA simply supposes that the topic distribution follows a Dirichlet prior [29]. This approach helps us to cluster topics across large data sets of news.

The application of this method to business problems is still very limited. Aral et al. [30] use LDA to extract common topics among 2,397 stock recommendations; Creamer et al. [7] apply LDA to identify common topics in a corporate network and use it to forecast return; Bao and Datta [31] use an extended version of LDA topic model to evaluate the effect of risk disclosures in 10-K forms on the risk perception of investors, and Xie et al. [32] test several NLP methods such as bag of words, LDA, and semantic frames for stock price prediction.

4) Centrality and Connectedness: Some of the centrality indicators that characterize an undirected graph $G(V, E)$ where $V = v_1, v_2, ..., v_n$ is the set of vertices, E is the set of edges, and e_{ij} is the edge between vertices v_i and v_j:

- Degree centrality $D_c(v_i) \doteq \sum_j a_{ij}$ where a_{ij} is an element of the adjacent matrix A of G and n is the number of vertices in G. Degree centrality is simply the sum of the edges of a vertex v_i.
- Betweenness centrality $B_c(v_i) \doteq \sum_i \sum_j \frac{g_{kij}}{g_{kj}}$. This is the proportion of all geodesic distances of all other vertices that include vertex v_i where g_{kij} is the number of geodesic (shortest) paths between vertices k and j that include vertex i, and g_{kj} is the number of geodesic paths between k and j [33].
- Eigenvector centrality $EV_c(v_i) \doteq \frac{\sum_{j=1}^{m} a_{ij} EV_c(v_j)}{\lambda}$. This is an indicator of the importance of a node v_i based on the sum of the centralities of its m neighbors. The matrix version is $AEV_c = \lambda EV_c$ where EV_c is the eigenvector of the largest eigenvalue λ of the adjacent matrix A [34, 35].

This research also uses the Krackhardt connectedness score [36] as a density measure of the digraph $C(V, E)$. Krackhardt connectedness is the fraction of all pairs of vertices v_i and v_j that has an undirected path between them in relation to all possible paths so that each vertex can reach any other vertex in C: $connectedness \doteq 1 - \frac{V}{\frac{N(N-1)}{2}}$

The range of the connectedness score is from zero for the null graph to one for the weakly connected graph.

B. Data

This study utilizes the daily equity prices of the components of the STOXX 50 index, and the STOXX 50 index itself provided by Thomson Reuters Tick History database, and the machine readable news offered by Thomson Reuters NewsScope for the period 2005-6/2011.

STOXX 50 includes the top 50 European companies by level of capitalization.

C. Research design

In this paper I propose an algorithm called CorpNetRisk (see Figures 1 and 2) that formalizes the procedure introduced in this section. I expect that companies that have much news in common may also behave alike. Therefore, the centrality indicators of each asset of a network based on common topics may represent the importance that a company has in a specific market. Additionally, the centrality of a company might also be associated with its volatility, and therefore VaR. Companies with high centrality may also be very stable and profitable as they play a central role in a network, so they might be reliable for the rest of the network; however, unstable companies with low and high returns may have periods that have very few common topics with other companies and other periods where they are highly connected or with high centrality values. The association between centrality and volatility might become more important during periods of crisis as market and systemic risk increases, and the number of news and common topics among companies may also increase.

I calculated daily volatility using a GARCH(1,1) model with a one year moving window (252 trading days) to control for autocorrelation and heteroscedasticity. The GARCH(1,1) volatility is the main input to calculate 99.9% VaR as the prime indicator of market risk for the large-cap European equity market (see Section II). Considering that the VaR used in this research is a linear transformation of volatility, in practical terms the results of the causality analysis applies to both volatility and VaR.

The main textual inputs are the news from Thomson Reuters. After eliminating the most common and redundant words using a stop word list, I extracted their common topics identified by the topic model methodology or LDA [28] (Fig. 1, steps 1-2, and Fig. 2), and matched them with the companies associated with every news. As a result of this process, I generated an asset-topic matrix which is a frequency table of the number of topics by company (Fig. 1, step 3).

With the asset-topics matrix, I built a dynamic social network based on a sequence of daily corporate news network from January 2005 to June 2011 (1,665 days). In this network, the nodes are companies, and the edges are the number of news items on the same topic by every pair of companies (Fig. 1, step 4). In total, I used 23,831,564 news items, and every company or node had associated 286 news items in average per day. However, the number of unique news items per company is smaller, about 30 daily news per company, as every news has several topics and is associated to several companies.

The companies used to build the news networks are 46 out of the 50 components of the STOXX 50 index in the third quarter of 2011. I eliminated 4 companies with incomplete information. This approach follows the tradition of using proximity-based networks for financial prediction as proposed by Mantegna [37].

I conducted a longitudinal analysis using the dynamic social network and calculated betweenness centrality, degree centrality, and eigenvector centrality [34, 35] for each node of the daily network (Fig. 1, step 5).

Using a daily moving window based on the previous month (22 training days), I evaluated if betweenness centrality, degree centrality, and eigenvector centrality of the constituents of the STOXX 50 index had a causal relationship to the next period VaR of each stock (Fig. 1, step 6). Based on these results, I calculated the CCI as the proportion of companies that show significant dependence between each centrality measure (betweenness centrality (CCI BC), degree centrality (CCI DC) and eigenvector centrality (CCI EVC)), and the next period VaR of each stock using the distance correlation t-test of independence (Fig. 1, step 7). Finally, I evaluated the causality of the seven lags of each CCI and their combined effect on the STOXX 50 index VaR, and selected the most relevant indicators (Fig. 1, step 8). In total the following parameters are used for the causality analysis of VaR:

1) Centrality: betweenness (BC), degree (DC) and eigenvector (EVC) (Table III-a).
2) Component Causality Index (CCI): betweenness centrality, degree centrality, and eigenvector centrality (Table III-b).
3) Combined: CCI: the three CCIs tested simultaneously (Table III-c).
4) Cent: the three centrality measures (betweenness, degree, and eigenvector) tested simultaneously (Table III-c).
5) CCIs-Cent: the three CCI s and the three centrality measures tested simultaneously (Table III-c).

I conducted two groups of causality tests to evaluate the network effect on volatility: 1) a Granger linear causality test, and 2) a linear and nonlinear causality test using the distance correlation for the following periods: before the financial crisis of the late 2000s (2005-06), the financial crisis (1/2007-3/2009) and the recovery period (4/2009-6/2011). The calculations for these two tests and the centrality measures were obtained using the vars, energy, and sna packages for R respectively.[1] I also applied the Bonferroni correction for the p-values of the distance correlation test, Granger causality test, and the comparison between these two tests.

IV. Results

Volatility, the centrality measures and their CCIs are stationary series according to the Augmented Dickey Fuller test (Table I-a). Therefore, the distance correlation test for independence can be used with these variables. The STOXX 50 log return series shows autocorrelation

[1] Information about R can be found at <http://cran.r-project.org>.

Input: Machine readable news and stock prices.

1) Preprocess news associated with selected assets: eliminate redundant items and delete stop words.
2) Cluster all the news using LDA. Every news is matched to at least one topic.
3) Build an asset-topics matrix. The element of this matrix is the number of news items that belong to a specific asset and that have been matched to a certain topic.
4) Build a longitudinal social network based on the asset-topics matrix where for each point in time the nodes are assets and the edges are the number of times that both assets have the same topic in common.
5) Calculate social network indicators (betweenness centrality, degree centrality, and eigenvector centrality) of the corporate news network.
6) Evaluate causality between social network indicators and VaR of each stock.
7) Calculate the component causality index as the proportion of assets of a particular index that have significant causal relationships between a social network indicator and their VaR over a given period.
8) Select leading indicator(s) for market risk according to the causality between the component causality index and the VaR of the system.

Output:
Leading indicator(s) for market risk.

Fig. 1: CorpNetRisk Algorithm

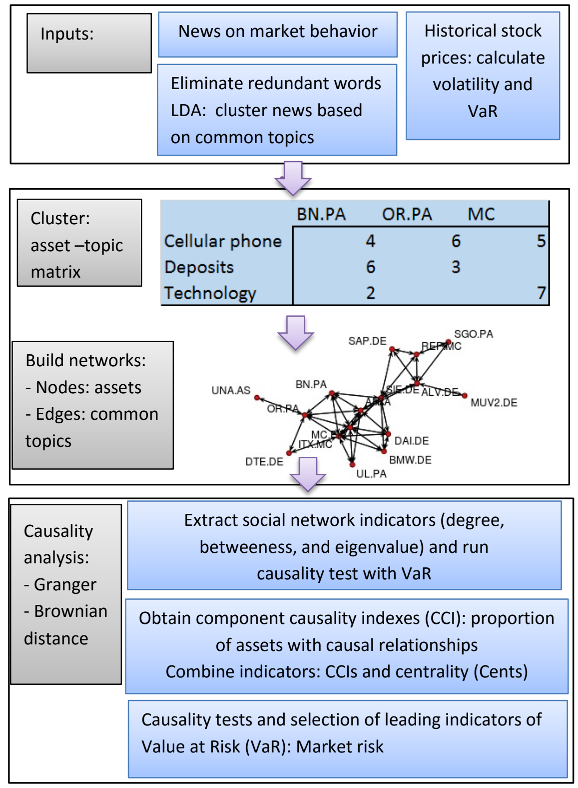


Fig. 2: CorpNetRisk Algorithm: social network and market risk

(Fig. 5-a), and autoregressive conditional heteroscedastic (ARCH) effect (Table I-b). Hence, a GARCH volatility model that takes into account this ARCH effect is adequate to obtain the STOXX 50 volatility.

The distance correlation test of independence captures a significantly larger number of causal relationships than the Granger causality test (Table II). A plausible explanation is that the relationship between centrality indicators and VaR is non-linear. Therefore the Granger causality test may underestimate it. However, the Granger causality analysis complements the Brownian distance correlation as it controls for the autoregressive effect of volatility. So, I will use both tests to study the CCIs calculated with the distance correlation test of independence as a leading indicator of the STOXX 50 index VaR.

Degree centrality and eigenvalue centrality are the most relevant individual measures that show significant relationships with STOXX 50 VaR during the first two periods of analysis according to the Brownian distance correlation (Table III-a). The three individual CCIs measures and the combination of these three measures (CCIs) improve the correlation and are significant during all the periods (Table III-b and c).The combination of all the centrality measures (Cent) or all the CCIs and centrality measures (CCIs-Cent) are relevant during the first two periods.

One of the main concerns of this analysis is that the results obtained are because of the volatility autocorrelation effect and their indirect effects on the explanatory variables. Granger causality isolates the contribution of each factor and also controls the autoregressive effect of the dependent variable, in this case, volatility or VaR. In this research, the Granger causality test confirms the main findings of the distance correlation test. It recognizes degree centrality and CCI betweenness centrality as the most relevant indicators during the complete period (Table III-a and b), and CCI degree centrality during the crisis period (2007-2009, Table III-b). Additionally, the three CCI measures alone and combined with the three centrality measures (CCIs-Cent) Granger cause the STOXX 50 VaR, especially during the first two periods of analysis (Table III-c).

According to Fig. 3, CCI degree centrality follows the changes of STOXX 50 VaR and the STOXX 50 index during the credit crisis period 2008-2009. CCI degree centrality shows major increases on 8/22/2008, 10/1/2008, 11/5/2008, 12/30/2008, and 1/23/2008. These peaks are followed by an increase of the STOXX 50 VaR and a fall of the STOXX 50 index. The most notorious peaks happen on 8/22/2008 and 1/23/2008 that anticipated the critical month of September 2008 with the precipitation of the U.S. financial crisis and March 6, 2009 which is the lowest point of the STOXX 50 during the period of analysis. Therefore, the most relevant series such as the CCI degree centrality or a combination of the CCIs can be used as leading market risk indicators, and could also be used as leading indicators for systemic risk. These results are also consistent with an increase on the return correlations of the STOXX 50 companies or the higher integration of the STOXX 50 corporate networks during the crisis period as observed in Fig. 4.

During crises, more news stories are generated because companies may be affected by external events or may take more actions either for strategic reasons or to protect them. As some of these events or actions may be common among companies with the same underlying risk factors, those companies may have more news with similar common topics. The fact that more companies share topics during periods of crisis leads to stronger connections among them, and it also increases the centrality indicators of the corporate news network. In this respect, Fig. 5 shows that the observations with the highest volatility also has a high level of connectedness. In particular, the 5% most and least volatile observations have average daily standard deviations of 0.038 and 0.005 and average Krackhart connectedness scores of 0.4 and 0.33 respectively.

In general, the causality analysis shows that the CCIs act as leading indicators of periods of higher volatility and VaR.

A. Network structure and risk

A final remaining question is if the topic model algorithm selected topics that capture different network dynamics. To explore this question, I use the community detection method proposed by Clauset et al. [38]. This method finds subgraphs or communities based on the greedy optimization of the modularity score. Modularity is a measure of the quality of a partition that evaluates if there are many links within communities and few links between communities. The main idea is that if the algorithm detects several communities when there are important differences in volatility, the structure of the network, and the topics that define the network structure, might be associated with volatility changes.

As the dataset has 1,665 daily networks, I only selected those networks that are associated with periods of maximum (November 4th., 2008) and minimum (January 3rd., 2006) standard deviation of the log return to illustrate the associated change of the network structure. I compared the communities discovered by country of residence, exchange, and economic sector of the companies that are part of each community.

The algorithm detects two communities for the network with the lowest average daily volatility and three for the network with the highest average daily volatility. There is not any major volatility difference among the communities of the same network, although there are important differences among the communities of a different network. The two communities of the low volatility network show similar distributions of economic sectors. The most important difference is that about half of the companies of the first and second communities are from Germany and Italy respectively (Table IV).

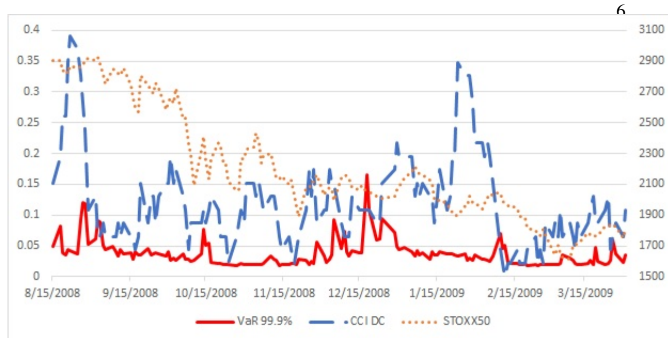


Fig. 3: Daily component causality index of degree centrality (CCI DC), Value at Risk at 99.9% confidence level, and STOXX 50 index (right axis).

The first, second and third community of the high volatility network are concentrated in the financial, industrial, and service sector respectively. For the second community, financial is the second most important sector. These communities also have a different distribution of countries of residence: a third part of the companies of the first community is from Italy, and about half of the companies of the second and third communities are from Germany, and France respectively. The importance of the industrial sector for the second community partially explains that the preferred country of residence for this community is Germany considering the relevance of the German industry in Europe.

This analysis illustrates how changes in the network structure might be associated with volatility changes. The importance of economic sectors and countries of residence in periods of high volatility might be due to a rise in the sector and country risk. As a result, the number of news associated with these aspects will also increase. Additionally, as the topics model methodology extracts common topics from the news, it might become easier to detect names of countries, cities or industrial terms that might be present in the news of several similar companies. The proposed method can also capture other trends associated with different risks. However, all of them combined have a final effect on detecting market risk.

V. Final Comments

This paper demonstrates that the corporate network structure defined by common topics of relevant news have a significant relationship with the next period market risk and volatility. The component causality index could also

	2005-06	2007-3/09	4/2009-11	2005-11
Log return	-8.88	-8.84	-9.13	-11.79
Volatility	-4.72	-5.51	-5.17	-8.80
Betweeness	-6.37	-6.63	-7.01	-7.29
Degree	-5.38	-5.58	-5.05	-9.01
Eigenvalue	-7.63	-5.95	-6.17	-9.53
CCI BC	-5.10	-5.20	-5.55	-8.76
CCI DC	-5.58	-6.19	-4.53	-8.08
CCI EVC	-5.30	-6.17	-6.12	-9.14

(a) ADF test of stationarity

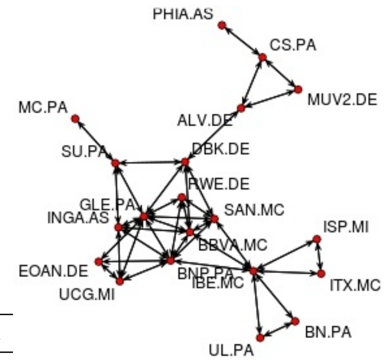
	2005-06	2007-3/09	4/2009-11	2005-11
a_t^2 of AR(1)	101.90	293.52	41.81	971.21
a_t^2 of GARCH(1,1)	16.05	7.16	1.08	8.30

(b) Ljung Box test of squared residuals (χ^2)

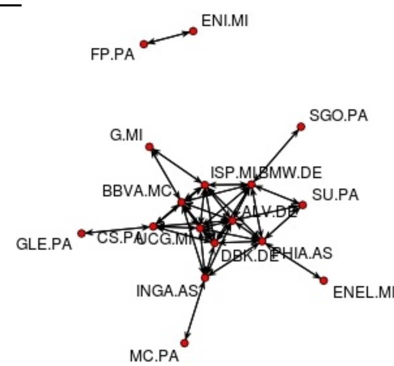
TABLE I: (a) t-statistic of the Augmented Dickey Fuller (ADF) unit-root test by period. The null hypothesis is that the series is non-stationary. (b) Chi-square of the Ljung Box test of squared residuals of AR(1) to test the ARCH effect of the STOXX 50 log return (1st. row) and of squared residuals of GARCH(1,1) (2nd. row) with seven lags. The null hypothesis is that there is no autocorrelation. All the series of the ADF test and the ARCH effect test have p-values $\leq$ 0.01. a_t^2 of GARCH(1,1) have p-values $>$ 0.01.

Method	Indicator	05-06	07-3/09	4/09-6/11	05-11
Distance Correlation	CCI BC	7%**	9%**	8%**	8%**
	CCI DC	11%**	10%**	12%**	11%**
	CCI EVC	10%**	10%**	11%**	10%**
Granger causality	CCI BC	8%	7%	6%	7%
	CCI DC	8%	7%	7%	7%
	CCI EVC	8%	7%	7%	7%

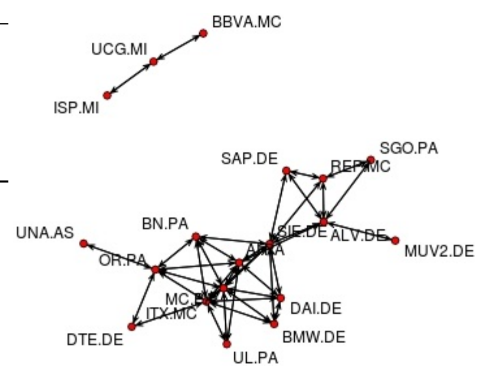
TABLE II: Component causality index (CCI) or proportion of STOXX 50 companies where centrality indicators have an effect on next period VaR based on daily data. CCI BC, CCI DC, and CCI EVC refer to CCI betweenness centrality, CCI degree centrality, and CCI eigenvalue centrality respectively. ANOVA shows a significant difference at 99% confidence level between distance and Granger causality across all the CCIs. **: p-value $\leq$ 0.01 with Bonferroni correction for the t-test mean difference between distance correlation and Granger causality.



(a) 2005-06



(b) 2007-09



(c) 2009-11

Fig. 4: Corporate networks based on the top twenty distance correlations among all the possible pairs of returns of the STOXX 50 companies. The nodes represent companies and the edges are the distance correlation between the pairs of assets return. The names of the nodes are the RICs or the companies' codes used by Thomson Reuters.

be used with different corporate networks such as trading networks in the energy sector. Although the trading activity is closely related to energy price movements, the behavior of the components of a particular market or system may have an impact on the risk of the system. In this respect, the proposed CCIs can also be part of a risk management model that includes the corporate news network effect and the main accounting, financial and economic variables to forecast market risk, and possibly systemic risk.

The broad impact of this research lies in the understanding of mechanisms of financial instability and risk as a result of a complex interaction of the dynamics of social groups and financial products. Problems of global financial instability are generally solved using short-term measures that limit the most evident effects of the crisis, not its causes. For instance, banning short selling positions is a typical tactic of economic authorities in periods of high financial instability (i.e. September 2008 and August 2011 in U.S.A. and Europe). However, the speculative nature of traders and their informational

	2005-06			2007-3/2009			4/2009-2011			2005-2011		
Lags	BC	DC	EVC	BC	DC	EVC	BC	DC	EVC	BC	DC	EVC
1	-0.31	5.05 **	1.73 *	1.34 .	2.58 ** †	5.89 **	0.09	0.40	0.24	1.13	0.14 †	0.81
2	0.37	8.28 ** $\mp$	3.83 **	2.36 **	2.94 **	4.47 **	-1.03	-0.67	0.47	0.42	-1.19 ‡	-0.16
3	1.26	9.14 ** $\mp$	2.68 **	0.30	0.67	3.67 **	-0.98	-0.49	-0.98	0.16	-0.40 ‡	-0.17
4	0.38	7.59 ** ‡	2.31 *	0.80	0.26	4.13 **	-0.51	-0.27	-0.97	-0.51	-0.66 ‡	-0.55
5	0.69	9.00 ** ‡	4.67 **	0.71	0.53	4.79 **	0.76 †	-0.42	-0.29	-0.33	-0.49 ‡	-0.07
6	-0.51	7.00 ** †	2.69 **	-0.32	2.45 **	3.90 **	0.84	-0.26	0.13	-0.69	-0.94 †	-1.01
7	-0.59	7.71 ** †	1.15	0.11	5.24 **	6.41 **	0.75	0.32	-0.30	-0.39	-0.84 †	-0.51

(a) Centrality Measures

	2005-06			2007-3/2009			4/2009-2011			2005-2011		
Lags	BC	DC	EVC	BC	DC	EVC	BC	DC	EVC	BC	DC	EVC
1	0.96	-0.53	4.92 **	2.94 **	15.61 ** $\mp$	3.26 **	2.67 **	4.59 **	1.49 .	2.63 **	4.77 **	8.84 **
2	6.06 ** ‡	0.19	5.63 **	3.46 **	16.15 ** $\mp$	2.30 *	3.37 **	4.20 **	1.41 .	7.28 ** ‡	6.62 **	8.66 **
3	9.19 ** ‡	1.25	6.39 **	2.84 **	18.86 ** $\mp$	2.94 **	4.63 **	4.37 **	0.67	7.77 ** †	7.46 **	8.70 **
4	10.76 ** †	1.13	7.71 **	3.97 **	19.05 ** ‡	3.35 **	5.72 **	4.17 **	1.28 .	9.12 **	10.53 **	10.12 **
5	8.77 **	1.17	9.08 **	2.77 **	17.26 ** ‡	4.38 **	5.63 **	3.82 **	1.68 *	5.93 **	9.65 **	11.89 **
6	9.67 **	1.07	8.05 **	2.73 **	19.57 ** †	6.50 **	7.45 **	2.19 *	3.06 **	8.94 ** †	10.30 **	14.15 **
7	9.20 **	2.20 *	10.31 **	3.85 **	24.88 ** ‡	6.27 **	7.49 **	1.87 *	2.26 *	10.06 **	13.43 **	15.58 **

(b) Component Causality Index (CCI)

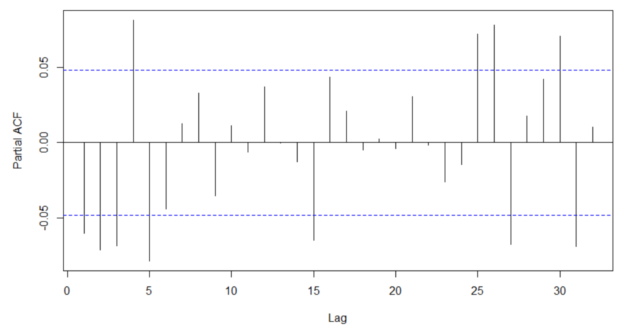
	2005-06			2007-3/2009			4/2009-2011			2005-2011		
	CCIs	Cent	CCIs-Cent	CCIs	Cent	CCIs-Cent	CCIs	Cent	CCIs-Cent	CCIs	Cent	CCIs-Cent
1	3.40 ** †	5.05 **	5.05 **	13.00 ** $\mp$	2.58 **	2.58 ** ‡	4.95 **	0.40	0.40	9.00 **	0.14	0.14
2	6.16 ** ‡	8.28 ** ‡	8.28 ** $\mp$	13.75 ** ‡	2.94 **	2.94 ** †	5.16 **	-0.67	-0.67	12.61 ** †	-1.19	-1.19 ‡
3	8.58 ** ‡	9.13 ** †	9.13 ** ‡	15.38 ** ‡	0.67	0.67 †	5.29 **	-0.49	-0.49	13.74 **	-0.40	-0.40
4	9.85 ** ‡	7.59 **	7.59 ** ‡	16.05 ** †	0.26	0.26	5.62 **	-0.27	-0.27	16.80 **	-0.66	-0.66
5	9.73 ** †	9.00 **	9.00 ** †	14.69 ** †	0.53	0.53	5.80 **	-0.42	-0.42	15.18 **	-0.49	-0.49
6	9.80 ** †	7.00 **	7.00 **	16.56 **	2.45 **	2.45 **	6.22 **	-0.26	-0.26	18.44 **	-0.94	-0.94
7	11.04 **	7.71 **	7.71 **	20.06 **	5.24 **	5.24 **	5.69 **	0.32	0.32	21.79 **	-0.84	-0.84

(c) Combined Component Causality Index (CCI)

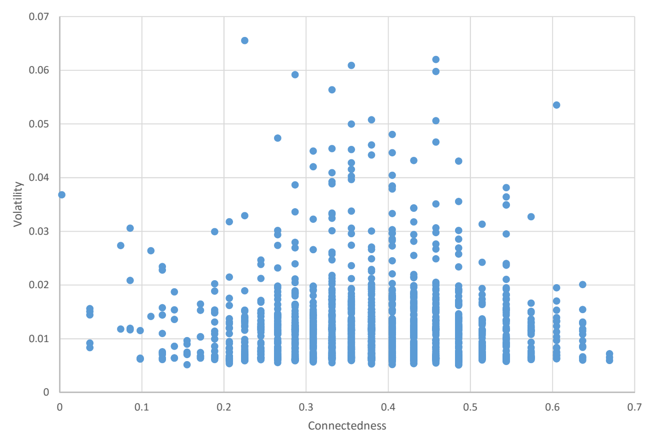
TABLE III: t-statistic of the distance correlation test of independence between centrality indicators, component causality index (CCI) and VaR for STOXX 50 index using daily data. In (a) BC, DC, and EVC refer to betweenness centrality, degree centrality, and eigenvalue centrality respectively, and in (b) refer to CCI betweenness centrality, CCI degree centrality, and CCI eigenvalue centrality respectively. In (c), CCIs refers to all CCI together (betweenness centrality, degree centrality, and eigenvalue centrality), Cent stands for all centrality indicators (betweenness, degree, and eigenvalue), and CCIs-Cent stands for the combination of all CCI and centrality measures. **, *, and .: p-value $\leq$ 0.01, 0.05 and 0.1 respectively with Bonferroni correction. This table also includes the p-values with Bonferroni correction for the Granger causality test applied to the same variables: $\mp$, ‡, and † indicate p-values $\leq$ 0.01, 0.05 and 0.1 respectively.

Regime	Com.	Volatility	Economic sectors				Country of residence						
			Res.	Industrial	Services	Financial	Germany	Spain	France	Ireland	Italy	Lux.	Neth.
High volatility	1	0.068		3	2	5	1	2	2	1	3		1
	2	0.078	2	7		5	6	2	4		2		
	3	0.067	2	2	4	1	2	1	5			1	
Low volatility	1	0.008		3	3	4	5	2	2	1			
	2	0.009	2	3	2	3	1	2	3		4		

TABLE IV: Community membership of corporate news network for different volatility regimes according to Clauset et al. [38]. Com., Lux. and Neth. refer to Communities, Luxembourg, and Netherlands respectively. Res. refers to Basic resources, oil and gas.



(a) PACF



(b) Krackhardt connectedness

Fig. 5: (a) Partial autocorrelation of the STOXX 50 log return series, and (b) STOXX 50 daily volatility and Krackhardt connectedness for the period 2005-2011.

advantage in relation to the rest of the society is a very difficult problem to solve with short-term measures.[2] The famous 2008 insider trading scandal of Goldman Sachs by Rajat Gupta, its former director, and Raj Rajaratnam, the hedge fund manager, shows the level of sophistication of this practice among investment bankers and traders and how difficult it is to probe this practice in court. The approach presented in this paper could be used to uncover these critical cases as it integrates social and economic network analysis with qualitative and quantitative variables that capture communicative and financial activity within these networks.

The techniques developed here are likely to apply outside of the financial analysis. Because many facets of society are relational, and because short text messages abound, inferred network structure and inferred sentiment might be used to predict other phenomena such as political events, industry growth, energy usage, or social trends.

REFERENCES

[1] F. Allen and D. Gale, "Financial contagion," *The Journal of Political Economy*, vol. 108, no. 1, pp. 1–33, 2000.

[2][39] find no evidence that the naked short selling ban of August 2008 altered the fundamental stock behavior.

[2] A. Dasgupta, "Financial contagion through capital connections: A model of the origin and spread of financial panics," *Journal of the European Economic Association*, vol. 2, no. 6, pp. 1049–1084, 2004.

[3] F. Allen and A. Babus, "Networks in finance," in *The Network Challenge: Strategy, Profit, and Risk in an Interlinked World*, P. R. Kleindorfer, Y. Wind, and R. E. Gunther, Eds. Upper Saddle River, NJ: Wharton School Publishing, 2009.

[4] S. Battiston, D. Delli Gatti, M. Gallegati, B. C. Greenwald, and J. E. Stiglitz, "Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk," *Journal of Economic Dynamics and Control*, vol. 36, no. 8, pp. 1121–1141, 2012.

[5] V. Dhar, T. Geva, G. Oestreicher-Singer, and A. Sundararajan, "Prediction in economic networks," *Information Systems Research*, vol. 25, no. 2, pp. 264–284, 2014.

[6] G. Creamer and S. Stolfo, "A link mining algorithm for earnings forecast and trading," *Data Mining and Knowledge Discovery*, vol. 18, no. 3, pp. 419–445, 2009.

[7] G. G. Creamer, Y. Ren, and J. V. Nickerson, "Impact of dynamic corporate news networks on asset return and volatility," *IEEE International Conference on Social Computing*, pp. 809–814, 2013.

[8] L. Adamic, C. Brunetti, J. Harris, and A. Kirilenko, "Information flow in trading networks," 2009, paper presented at 1st. Workshop on Information in Networks, NYU, NY.

[9] M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, pp. 535–559, 2012.

[10] Z. Zheng, B. Podobnik, L. Feng, and B. Li, "Changes in cross-correlations as an indicator for systemic risk," *Scientific Reports*, vol. 2, no. 888, pp. 1–8, 2012.

[11] N. Casnici, P. Dondio, R. Casarin, and F. Squazzoni, "Decrypting financial markets through e-joint attention efforts: On-line adaptive networks of investors in periods of market uncertainty," *PLOS ONE*, vol. 10, no. 8, pp. 1–15, 2015.

[12] D. Bisias, M. Flood, A. W. Lo, and S. Valavanis, "A survey of systemic risk analytics," 2012, Office of Financial Research, Working Paper No. 0001.

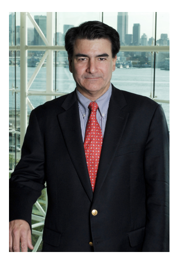
[13] R. J. Shiller, *Irrational Exuberance*, 3rd ed. Princeton, NJ: Princeton University Press, 2015.

[14] R. Chatterjee, *Practical Methods of Financial Engineering and Risk Management*. Apress-Springer, 2014.

[15] D. Ruppert and D. S. Matteson, *Statistics and Data Analysis for Financial Engineering with R examples*, 2nd ed. New York: Springer, 2015.

[16] R. Rockafellar and S. Uryasev, "Optimization of

conditional value-at-risk," *Journal of Risk*, vol. 2, pp. 21–41, 2000.
[17] P. Jorion, *Financial Risk Manager Handbook*, 6th ed. Hoboken, NJ: Wiley, 2011.
[18] J. Longerstaey and L. More, *Introduction to RiskMetrics*. Morgan Guaranty Trust Company, 1995.
[19] S. Goyal, *Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press, 2002.
[20] M. Jackson, *Social and Economic Networks*. Princeton, NJ: Princeton University Press, 2008.
[21] ——, "The study of social networks in economics," in *The Missing Links: Formation and Decay of Economic Networks*, J. Podolny and J. Rauch, Eds. New York: Russell Sage Foundation, 2007.
[22] F. Allen and D. Gale, *Understanding Financial Crises*. Oxford: Oxford University Press, 2007.
[23] C. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.
[24] M. Ding, Y. Chen, and S. L. Bressler, "Granger causality: Basic theory and application to neuroscience," in *Handbook of Time Series Analysis*, B. Schelter, M. Winterhalder, and J. Timmer, Eds. Wienheim: Wiley, 2006, pp. 437–460.
[25] G. J. Székely and M. L. Rizzo, "Brownian distance covariance," *The Annals of Applied Statistics*, vol. 3, no. 4, pp. 1236–1265, 2009.
[26] ——, "The distance correlation t-test of independence in high dimension," *Journal of Multivariate Analysis*, vol. 117, pp. 193–213, 2013.
[27] T. Tsay, *Analysis of Financial Time Series*, 3rd ed. Hoboken, NJ: Wiley, 2010.
[28] M. Steyvers and T. Griffiths, "Probabilistic topic models," in *Latent Semantic Analysis: A Road to Meaning*, T. Landauer, D. McNamara, S. Dennis, and W. Kintsch, Eds. Mahwah, NJ: Erlbaum, 2007.
[29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Mar. 2003.
[30] S. Aral, P. G. Ipeirotis, and S. Taylor, "Content and context: Identifying the impact of qualitative information on consumer choice," in *32nd International Conference on Information Systems (AIS, Shanghai)*, pp. 511–525, 2011.
[31] Y. Bao and A. Datta, "Simultaneously discovering and quantifying risk types from textual risk disclosures," *Management Science*, vol. 60, no. 6, pp. 1371–1391, 2014.
[32] B. Xie, R. J. Passonneau, L. Wu, and G. Creamer, "Semantic frames to predict stock price movement," in *51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria: Association for Computational Linguistics, 2013.
[33] L. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–39, 1979.
[34] P. Bonacich, "Factoring and weighting approaches to clique identification," *Journal of Mathematical Sociology*, vol. 2, pp. 113–120, 1972.
[35] ——, "Some unique properties of eigenvector centrality," *Social Networks*, vol. 29, no. 4, pp. 555–564, 2007.
[36] D. Krackhardt, "Graph theoretical dimensions of informal organizations," in *Computational Organization Theory*, K. M. Carley and M. J. Prietula, Eds. Hillsdale, NJ: Lawrence Erlbaum and Associates, 1994.
[37] R. N. Mantegna, "Hierarchical structure in financial markets," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 11, no. 1, pp. 193–197, 1999.
[38] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, p. 066111, Dec 2004.
[39] C. Ulibarri, I. Florescu, and J. Eidsath, "Regulating noisy short-selling of troubled firms?" *Journal of Financial Economic Policy*, vol. 1, no. 3, pp. 227–245, 2009.



Germán G. Creamer is an Associate Professor of quantitative finance and business analytics at Stevens Institute of Technology and an adjunct associate professor at Columbia University. Dr. Creamer has been a senior manager in the Risk, Information and Banking Division in American Express where he worked in the enterprise-wide risk management and the information management groups. He has taught at Tulane University, and in several leading Latin American business schools. Dr. Creamer has been economic advisor to the president of Ecuador and the government of Equatorial Guinnea. He has also consulted for several hedge funds and international organizations such as United Nations, World Bank, and US Agency for International Development. His current area of research is on trading systems, risk management and financial analytics. Dr. Creamer has a PhD in Computer Science (specialized on computational finance), a MSc in Financial Engineering, both from Columbia University, and a PhD in Economics from the University of Notre Dame.