NETWORK ANALYSIS OF RETURNS AND VOLUME TRADING IN STOCK MARKETS: THE EURO STOXX CASE

Juan Gabriel Brida
Universidad de la República, Uruguay. Email: JuanGabriel.Brida@unibz.it

David Matesanz
Universidad de Oviedo, Spain. Email: matesanzdavid@uniovi.es

Maria Nela Seijas
Universidad de la República, Uruguay. Email: mseijas@ccea.com.uy

Abstract: This study applies network analysis to analyze the structure of the Euro Stoxx market during the long period from 2002 up to 2014. The paper generalizes previous research on stock market networks by including asset returns and volume trading as the main variables to study the financial market. A multidimensional generalization of the minimal spanning tree (MST) concept is introduced, by adding the role of trading volume to the traditional approach which only includes price returns. Additionally, we use symbolization methods to the raw data to study the behaviour of the market structure in different, *normal* and *critical*, situations. The hierarchical organization of the network is derived, and the MST for different sub-periods of 2002-2014 is created to illustrate how the structure of the market evolves over time. From the structural topologies of these trees, different clusters of companies are identified and analyzed according to their geographical and economic links. Two important results are achieved. Firstly, as other studies have highlighted, at the time of the financial crisis after 2008 the network becomes a more centralized one. Secondly and most important, during our second period of analysis, 2008-2014, we observe that hierarchy becomes more country-specific where different sub-clusters of stocks belonging to France, Germany, Spain or Italy are found apart from their business sector agrupation. This result may suggest that during this period of time financial investors seem to be worried most about country specific economic circumstances.

JEL codes: C45, F36, G12

Keywords: Correlation networks, symbolization methods, Minimum spanning tree, Ultrametric hierarchical tree, Taxonomy, financial crisis, Eurostoxx market

Research highlights

► the study proposes a hierarchical organization of the Eurostoxx market
► Volume and prices of assets are used jointly to analyze the structure of the market
► Symbolization methods are applied to analyze market behaviour in *normal* and *critical* situations
► a country stock organization is found during the financial crisis period aside from the traditional sectorial network organization
► the network becomes star-shaped when analyzing critical market conditions during period 2008-2014, later to financial crisis

1. Introduction

Asset returns and volume trading are the main variables to study stock markets and investors and academics alike have given considerable attention to the price-volume relationship over the past two decades. The importance of considering the volumes to analyze stock prices movements can be considered a well-accepted practice in the financial area. However, when we look at the scientific production in this field, we can only find few works including volume and price variations for stock assessment purposes (e. g., Clark, 1973; Morgan, 1976; Karpoff, 1987; Suominen, 2001; Podobnik et al., 2009). Souminen (2001) points out that, despite that academic literature in finance contains few theoretical papers on the role of trading volume, many practitioners use that information. Karpoff (1987) surveys previous research on two stylized facts. At first, as in an old Wall Street adage that says "It takes Volume to make prices move". Some studies have attempted to establish the empirical and theoretical structure of the relationship between volume and stock price change. The connection was introduced by Osborne (1959) in the literature and later developed by Granger and Morgenstein (1963), Godfrey, Granger and Morgenstein (1964), Crouch (1970), Clark (1973), Copeland (1976), Morgan (1976), Westerfield (1977), Cornell (1981), Jennings et al. (1981), Tauchen and Pitts (1983), Harris (1983), Saatcioglu and Starks (1998) and Assogbavi and Osagie (2006), among others. However, in spite of many studies that have been made in this area, we still do not have a unified model that aggregates traded volume and price variations for stock prices assessment purposes, especially for long-term investment horizon.

Another familiar adage says that, "Volume is relatively heavy in bull markets and light in bear markets". Karpoff remarks that, in equity markets there is evidence of positive relationship between volume and price change (see Morgan (1976), Jain and Joh (1988), Rogalski (1978) and Harris, (1986), among others). Karpoff concludes that even if these two empirical findings seem to set up a contradiction, it could be explained by an asymmetric volume-price change relationship, indicating that the relationship is fundamentally different for positive and negative price changes. Since volume trading seems to carry important information to the market, we aim to describe the Euro Stoxx 50 market, introducing a methodology which embodies information supplied not only by returns but also by volume trading. Lo and Wang (2000) assert that, if price and quantity are fundamental in any theory of market interactions, the importance of trading volume in modeling asset markets is clear. However, most of the models of asset markets have focused on the behavior of returns, giving far less attention to trading volume. It is well known that oscillations of stock prices are highly inter-coupled with strong correlations with the different business sectors and industries to which the stocks belong. Complex networks provide a very general framework, based on the concepts of statistical physics, for studying systems with large numbers of interacting agents. These networks have been able to successfully describe the topological properties and characteristics of many economic systems. Recently, analyzes based on network models have been proposed for studying the correlations of stock prices, including the works of Mantegna (1999), Mantegna and Stanley (1999), Vandewalle et al. (2001), Bonanno et al. (2001), Bonnanno et al. (2003), Onnela et al. (2003). Bonanno et al. (2004), Tumminello et al. (2005), Tse et al. (2008), Chi et al. (2010), Keskin et al. (2011), Lyócsa et al. (2012), Oatley et al. (2013), Kocakaplan et al. (2013), Heiberge (2014) and Birch et al. (2015).

The typical approach includes a process of finding a distance based on the correlation between each pair of time series of stock prices, and a subsequent procedure of constructing a network that connects the individual stocks based on the levels of correlation. In particular, the way of describing the structure of a market involves the construction of Minimal Spanning Trees (MST) and Hierarchical Trees (HT). These graphs show the interconnection among the firms, detecting clusters and taxonomic relationships in a financial market. Almost all the studies included in this line of research utilize a notion of distance based on the Pearson correlation coefficient as a function to measure the similarity between two time series.

In this paper we introduce an alternative metric to the Pearson correlation coefficient to obtain other MSTs and HTs. These distances are based on symbolic methods (see Daw et al. (2003); De Polsi et al., (2013); Sorrentino et al. (2014); Makkiet al. (2015)) where symbolization is constructed using an economic-statistical criteria. The paper shows that these tools can also be applied to describe the market topology in industrial networks. In fact, as remarked by (Halinen and Tornroos, 2005), it is necessary to study networks of inter-firms relationships in order to understand the functioning of industrial markets. They remark that there is still a great need to develop concepts and methodologies in network research. Hakansson and Ford (2002) also highlight that companies within a network are not free to act according to their own aims. They do not operate in isolation from others, or in response to some generalized environment. Instead, each company's considerations and actions can only be fully understood within a structure of individually significant counterparts and relationships. Some works focus on market topology (see (Burt and Carlton, 1989) and (Souma et al., 2006)) or spatial clustering, searching geographical industrial clusters (see (Feser and Sweeny, 2000) and (Henderson et al., 2002)). In this paper, the flexibility of the methodology is revealed, permitting to analyze market topology and dynamics in critical situations. In fact by changing the thresholds of the partition and focusing on extreme values the methodology allows studying the structure of the market in critical situations. This information can be useful when managing risk, in analysis of financial crashes and Value at Risk (see (Holton, 2003), (Jorion, 1997)).

This paper follows the Symbolic Time Series approach for clustering introduced by Mantegna (1999), Brida and Risso (2007; 2009), Brida et al. (2009), Gullapalli and Carley (2013) and Abbasi and Loun (2014). This symbolic method gives more flexibility to the analysis, but only applies for one dimensional time series, losing the possibility of embodying information from the volume trading. In this paper we propose a multidimensional generalization of the previous methods. By means of this generalization we firstly aim to describe the topology and hierarchy of the European stock market during the long period from 2002 up to 2014 both considering normal and critical stages[1].Secondly, we focus on the sub-period from 2008 to 2014 to analyze the effects of the global financial crisis on this topology and hierarchy.

[1] One of the main strengths of symbolization methods relies in the fact that it is not necessary to split the database in normal and critical periods to separately analyze these two economically different stages.

The paper is organized as follows. Section II explains the methodology. In section III we describe data and analyze the hierarchical structure of the European stock market. Section IV concludes the paper.

2. Multidimensional Symbolic based Minimal Spanning Tree

MST and HT are useful tools represented by visual nets, showing the most relevant connections and interactions in the stock market. To obtain these graph representations we applied a metric, computing all the distances between companies. Assume we have a multidimensional time series with real values, as follows:

$$\{Xi\}_{t=1}^{t=T} = \left\{ \begin{pmatrix} x_{i1} \\ y_{i1} \\ . \\ z_{i1} \end{pmatrix}, \begin{pmatrix} x_{i2} \\ y_{i2} \\ . \\ z_{i2} \end{pmatrix}, \dots, \begin{pmatrix} x_{iT} \\ y_{iT} \\ . \\ z_{iT} \end{pmatrix} \right\} \tag{1}$$

We can convert into a one-dimensional symbolic space S, by defining a determined threshold in the multidimensional space R^n , defined by the series x, obtaining the following coded symbolic time series for each company i:

$$\{s_{i1}, s_{i2}, \dots\dots s_{iT}\} \tag{2}$$

Once this symbolic time series is obtained for each firm, we can compute the distance between the companies, according to the following distance function:

$$d_0(s_i, s_j) = \sqrt{\sum_{t=1}^{t=T} (s_{it} - s_{jt})^2} \tag{3}$$

Note that $\{s_{i1}, s_{i2}, \dots\dots, s_{iT}\}$ and $\{s_{j1}, s_{j2}, \dots\dots, s_{jT}\}$ are two symbolic sequences for companies i and j respectively. When all the distances are computed, we can derive the structure of the financial market by constructing the MST and HT. The MST is progressively constructed by linking all the time series together in a graph characterized by a minimal distance between the firms. In words, we rank the obtained distances from the smallest to the largest and start linking the two firms with the shortest distance. In the second step, we take the second smallest distance in the rank, linking the corresponding pair of companies. We proceed linking all the companies according to the rank of distance until obtaining a graph connecting all the firms in the market, applying Kruskal algorithm. The MST permits to obtain the subdominant ultrametric distances $d^<$ permitting to construct the HT. This distance, $d^<$ (i,j) between i and j is the maximum value of any Euclidean distance d(l,m) detected by moving in single steps from i to j through the shortest path connecting i and j in the MST(see Ramal et al. (1986), for a definition of ultrametricity). Based on Euclidean distance, a complete method of hierarchical clustering was used to classify Euro Stoxx stocks into similar clusters. To choose the proper number of clusters, pseudo F, pseudo t-squared and R-Squared were used as detention measures.

3. Structure of the Euro Stoxx 50 Index

The Euro Stoxx 50 Index offers exposure to fifty of the largest stocks in the Euro zone. Considered as Europe's leading Blue-chip index for the Eurozone, this index provides a Blue-chip representation of super-sector leaders in the Eurozone. The index covers fifty stocks from seven Eurozone countries: Belgium, Finland, France, Germany, Italy, the Netherlands and Spain. The index is free float market capitalization weighted. A regular annual review may affect its composition, determining additions and deletions of blue-chip stocks from the index. In this paper we apply daily data from the Euro Stoxx 50 index for the period January 2, 2002 to December 19, 2014, collecting data of trading volume in shares and asset returns for fifty companies[2]. Table 1 shows the companies composing the Euro Stoxx 50 Index, indicating its name and acronym, business sector and country of origin, by its ISO codes.

Table 1 – Euro Stoxx 50 composition

No.	Acronym	Company	Business sector	Country
1	AI.FP	AIR LIQUIDE	Chemicals	FR
2	ALV.GY	ALLIANZ	Insurance	DE
3	ABI.BB	ANHEUSER-BUSCH INBEV	Food & Beverages	BE
4	ASML.NA	ASML HOLDING NV	Technology	NL
5	G.IM	ASSICURAZIONI GENERALI	Insurance	IT
6	CS.FP	AXA	Insurance	FR
7	BBVA.SQ	BCO BILBAO VIZCAYA ARGENTARIA	Banks	ES
8	SAN.SQ	BCO SANTANDER	Banks	ES
9	BAS.GY	BASF	Chemicals	DE
10	AIR.FP	AIRBUS GROUP NV	Aerospace & Defense	FR
11	BAYN.GY	BAYER	Chemicals	DE
12	BMW.GY	BMW	Automobiles & Parts	DE
13	BNP.FP	BNP PARIBAS	Banks	FR
14	CA.FP	CARREFOUR	Retail	FR
15	SGO.FP	SAINT GOBAIN	Construction & Materials	FR
16	DAI.GY	DAIMLER	Automobiles & Parts	DE
17	BN.FP	DANONE	Food & Beverages	FR
18	DBK.GY	DEUTSCHE BANK	Banks	DE
19	DPW.GY	DEUTSCHE POST	Industrial Goods & Services	DE
20	DTE.GY	DEUTSCHE TELEKOM AG	Telecommunications	DE
21	EOAN.GY	EON AG	Utilities	DE
22	ENEL.IM	ENEL SPA	Utilities	IT
23	ENI.IM	ENI SPA	Oil & Gas	IT
24	EI.FP	ESSILOR INTERNATIONAL SA	Healthcare	FR
25	IBE.SQ	IBERDROLA	Utilities	ES
26	ITX.SQ	INDUSTRIA DE DISENO TEXTIL SA	Retail	ES
27	INGA.NA	ING GRP	Banks	NL
28	ISP.IM	INTESA SANPAOLO	Banks	IT
29	PHIA.NA	PHILIPS	Industrial Goods & Services	NL

[2] GSZ FP Equity Shares and UL NA Equity, with fewer observations than the rest of stocks, were subsequently incorporated into the analysis for sub-period 2008-2014 but didn´t modify the results significantly.

30	OR.FP	L'OREAL	Personal & Household Goods	FR
31	MC.FP	LVMH MOET HENNESSY	Personal & Household Goods	FR
32	MUV2.GY	MUENCHENER RUECK	Insurance	DE
33	NOK1V.FH	NOKIA	Technology	FI
34	ORA.FP	ORANGE	Telecommunications	FR
35	REP.SQ	REPSOL	Oil & Gas	ES
36	RWE.GY	RWE	Utilities	DE
37	SAN.FP	SANOFI	Healthcare	FR
38	SAP.GY	SAP	Technology	DE
39	SU.FP	SCHNEIDER ELECTRIC	Industrial Goods & Services	FR
40	SIE.GY	SIEMENS	Industrial Goods & Services	DE
41	GLE.FP	GRP SOCIETE GENERALE	Banks	FR
42	TEF.SQ	TELEFONICA	Telecommunications	ES
43	FP.FP	TOTAL	Oil & Gas	FR
44	UCG.IM	UNICREDIT	Banks	IT
45	UNA.NA	UNILEVER NV	Personal & Household Goods	NL
46	DG.FP	VINCI	Construction & Materials	FR
47	VIV.FP	VIVENDI	Media	FR
48	VOW3.GY	VOLKSWAGEN PREF	Automobiles & Parts	DE
49	GSZ.FP	GDF SUEZ	Utilities	FR
50	UL.NA	UNIBAIL-RODAMCO	Real Estate	FR

Source: Bloomberg

The daily return on assets is calculated as the difference between the logarithms of the closing prices for the company i at time t and time t-1 while the daily trading volume is taken in logarithmic terms. Data of closing price and trading volume have been downloaded from the Bloomberg database. All data are valued in euros. Table 2 shows the average, maximum and minimum values of daily stock returns in the whole period analyzed, as well as sub-period 2002-2007 and 2008-2014, which can be considered as a phase of pre and post-crisis, respectively.

Table 2 – Summary statistics

		2002-14			2002-07			2008-14		
No.	Stock	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean	Max.
1	AI.FP	-25,54%	0,07%	26,26%	-20,23%	0,10%	20,01%	-25,54%	0,04%	26,26%
2	ALV.GY	-45,69%	-0,05%	70,75%	-34,08%	-0,08%	36,19%	-45,69%	-0,03%	70,75%
3	ABI.BB	-69,48%	0,13%	49,42%	-32,22%	0,11%	22,15%	-69,48%	0,14%	49,42%
4	ASML.NA	-51,89%	0,12%	65,37%	-51,89%	0,03%	65,37%	-28,17%	0,20%	38,65%
5	G.IM	-29,25%	-0,04%	39,96%	-22,28%	0,03%	22,48%	-29,25%	-0,10%	39,96%

6	CS.FP	-67,56%	-0,01%	66,24%	-39,93%	0,04%	52,25%	-67,56%	-0,06%	66,24%
7	BBVA.SQ	-49,78%	-0,05%	71,00%	-24,25%	0,04%	31,06%	-49,78%	-0,13%	71,00%
8	SAN.SQ	-45,46%	-0,02%	77,31%	-38,92%	0,10%	29,91%	-45,46%	-0,13%	77,31%
9	BAS.GY	-47,32%	0,10%	56,57%	-25,43%	0,17%	38,87%	-47,32%	0,03%	56,57%
10	AIR.FP	-103,92%	0,09%	35,05%	-103,92%	0,09%	35,05%	-41,71%	0,10%	35,03%
11	BAYN.GY	-47,36%	0,10%	105,10%	-47,36%	0,13%	105,10%	-34,19%	0,08%	38,98%
12	BMW.GY	-42,25%	0,07%	42,35%	-25,25%	0,01%	23,01%	-42,25%	0,11%	42,35%
13	BNP.FP	-60,14%	0,00%	60,66%	-33,59%	0,07%	37,73%	-60,14%	-0,07%	60,66%
14	CA.FP	-34,98%	-0,06%	27,24%	-28,13%	-4,33%	27,24%	-34,98%	-0,11%	25,47%
15	SGO.FP	-80,82%	-0,01%	50,05%	-80,82%	0,07%	41,15%	-50,75%	-0,09%	50,05%
16	DAI.GY	-50,36%	0,03%	61,98%	-22,69%	0,07%	28,48%	-50,36%	-0,01%	61,98%
17	BN.FP	-24,93%	0,04%	31,20%	-23,95%	0,12%	31,20%	-24,93%	-0,02%	23,56%
18	DBK.GY	-56,59%	-0,10%	66,68%	-25,67%	0,02%	28,03%	-56,59%	-0,20%	66,68%
19	DPW.GY	-59,97%	0,04%	51,23%	-24,95%	0,09%	21,50%	-59,97%	0,01%	51,23%
20	DTE.GY	-54,00%	-0,04%	47,65%	-54,00%	-0,05%	44,88%	-48,36%	-0,03%	47,65%
21	EOAN.GY	-41,56%	-0,03%	59,13%	-25,19%	0,20%	27,58%	-41,56%	-0,22%	59,13%
22	ENEL.IM	-40,89%	-0,04%	57,30%	-33,83%	0,08%	16,36%	-40,89%	-0,14%	57,30%
23	ENI.IM	-34,31%	0,00%	51,54%	-26,32%	0,12%	15,39%	-34,31%	-0,10%	51,54%
24	EI.FP	-25,57%	0,13%	37,83%	-19,53%	0,17%	25,67%	-25,57%	0,10%	37,83%
25	IBE.SQ	-45,96%	0,05%	59,24%	-15,47%	0,23%	50,63%	-45,96%	-0,10%	59,24%
26	ITX.SQ	-76,64%	0,15%	42,29%	-76,64%	0,12%	42,29%	-35,88%	0,17%	35,38%
27	INGA.NA	-111,61%	-0,07%	87,57%	-36,62%	-0,01%	52,07%	-111,61%	-0,12%	87,57%
28	ISP.IM	-66,17%	0,00%	67,02%	-30,23%	0,16%	32,26%	-66,17%	-0,14%	67,02%
29	PHIA.NA	-38,32%	-0,04%	42,80%	-37,12%	-0,02%	42,80%	-38,32%	-0,05%	36,36%
30	OR.FP	-26,81%	0,04%	39,25%	-20,86%	0,04%	23,05%	-26,81%	0,04%	39,25%
31	MC.FP	-34,35%	0,09%	35,00%	-19,45%	0,10%	31,47%	-34,35%	0,07%	35,00%
32	MUV2.GY	-39,82%	-0,05%	38,55%	-39,82%	-0,14%	36,97%	-33,42%	0,02%	38,55%
33	NOK1V.FH	-69,30%	-0,15%	109,52%	-66,45%	-0,01%	42,45%	-69,30%	-0,26%	109,52%
34	ORA.FP	-56,51%	-0,09%	73,99%	-56,51%	-0,08%	73,99%	-25,67%	-0,10%	37,25%
35	REP.SQ	-57,84%	0,00%	36,05%	-27,07%	0,09%	35,19%	-57,84%	-0,07%	36,05%
36	RWE.GY	-35,71%	-0,05%	47,12%	-21,75%	0,16%	25,99%	-35,71%	-0,22%	47,12%
37	SAN.FP	-35,17%	-0,02%	41,70%	-29,66%	-0,05%	26,07%	-35,17%	0,02%	41,70%
38	SAP.GY	-57,46%	0,04%	74,27%	-48,55%	0,01%	74,27%	-57,46%	0,06%	37,14%
39	SU.FP	-47,42%	0,07%	43,82%	-23,06%	0,11%	18,08%	-47,42%	0,04%	43,82%
40	SIE.GY	-61,84%	0,02%	68,32%	-23,24%	0,08%	32,48%	-61,84%	-0,04%	68,32%
41	GLE.FP	-55,76%	-0,05%	68,14%	-32,83%	0,08%	34,06%	-55,76%	-0,16%	68,14%
42	TEF.SQ	-33,46%	-0,01%	46,00%	-29,67%	0,12%	46,00%	-33,46%	-0,12%	39,85%
43	FP.FP	-31,13%	0,00%	41,14%	-20,67%	0,07%	23,00%	-31,13%	-0,06%	41,14%
44	UCG.IM	-65,62%	-0,15%	67,67%	-21,82%	0,05%	32,31%	-65,62%	-0,31%	67,67%
45	UNA.NA	-35,86%	0,03%	27,01%	-35,86%	0,03%	22,49%	-28,68%	0,03%	27,01%
46	DG.FP	-39,02%	0,09%	49,09%	-22,97%	0,22%	24,96%	-39,02%	-0,02%	49,09%
47	VIV.FP	-99,23%	-0,10%	64,35%	-99,23%	-0,13%	64,35%	-34,53%	-0,07%	41,91%
48	VOW3.GY	-56,01%	0,14%	50,18%	-28,42%	0,20%	23,94%	-56,01%	0,09%	50,18%
49	GSZ.FP	---	---	---	---	---	---	70,06%	-44,66%	-0,12%
50	UL.NA	---	---	---	---	---	---	21,74%	-22,85%	0,08%

Source: Own elaboration using R

Table 2 shows that, for most stocks, the maximum and minimum values for the entire period correspond to the extreme values for the sub-period 2008-2014, which a priori could be treated as the sub-period with greater volatility of returns. In order to obtain a partition in the bi-dimensional space (asset returns, volume trading), at first, we consider a kind of global return for each company *i*, given by the product between returns and volume trading at moment t: $R_i(t) = r_i(t) * V_i(t)$, being $r_i(t)$ defined as the difference between logarithm of prices for the company *i* at time *t* and *t*-1. For a series of size T, thus we can construct the empirical distribution $f^*(R_i)$ of these returns for company *i*.

According to Molgedey and Ebeling (2000) we should use small partitions with three pieces. We would like to study the market structure in both *normal* and *critical* situations. Therefore, for the *normal* situation we define three equally probable regions where empirical density cumulates 1/3 and 2/3 of the distribution. For the *critical* situation, we define thresholds where empirical distribution cumulates 15% and 85% of the distribution.

Each pair of return and volume trading takes a unique symbol according to the region they are in. Once the symbolization is complete we can compute all distances and construct the MST and the HT as previously explained. Note that the distance between two companies as in equation (3) represents how close the qualitative dynamics of asset returns and volume trading was during the period of study (see Brida et al. (2003) for a pedagogical exposition of this qualitative dynamics). This method could be useful in portfolio construction and management. Since Markowitz (1952) establishes that putting uncorrelated assets in a portfolio reduces the risk, many strategies to select assets have been applied using this idea.

Some authors such as Bouchaud and Potters (2000) suggest that what is relevant is if the firms are uncorrelated in the extreme events. According to them it is important to focus on the correlation at the tails of the distribution of the returns. We suggest that by defining the partitions weighting more the extreme situations (high negative and positive returns and high volume trading) we can obtain a tree showing the cluster of companies with the same dynamics. in critical situations, therefore it should be not convenient to put all the assets of the same cluster in the portfolio.

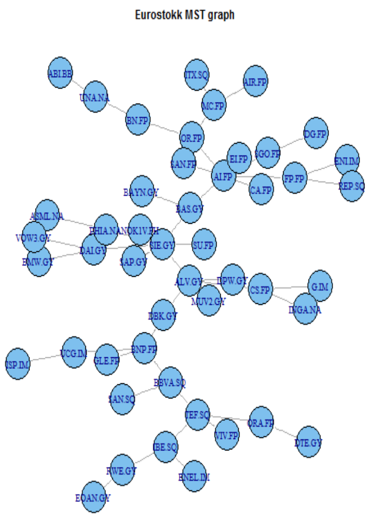
For the purpose of comparing the evolution of daily global returns for Euro Stoxx 50 stocks, it is useful to analyze different time periods. The data is then analyzed for the entire time-span set out and for two different sub-periods, 2002-07 and 2008-14[3]. All considered periods have been calculated for both, the *normal* and *critical* symbol codification. However, to gain an insight into the dynamics of global returns of this index, rolling windows, each corresponding to a calendar year, are investigated, from January, 2002 to December, 2014[4]. We distinguish eight cases, four corresponding to the *normal* situation (for the whole period, the first sub-period and the second sub-period, considering 48 and 50 stocks, respectively) and four for the *critical* situation, with similar order.

[3] In this paper, we show results just for the entire time sample and for the last period, 2008-2014. Results from 2002-2007 are available directly from the authors upon request.

[4] The last rolling window, covering data up to December, 19th 2014, is arranged to have the same number of observations that the rest of the windows, so as not to get any asymmetry in results.

3. Empirical results

The MST and HT obtained for the Euro Stoxx 50 case in a normal situation for the whole period (Case 1) are shown in Figure 1 and 2.



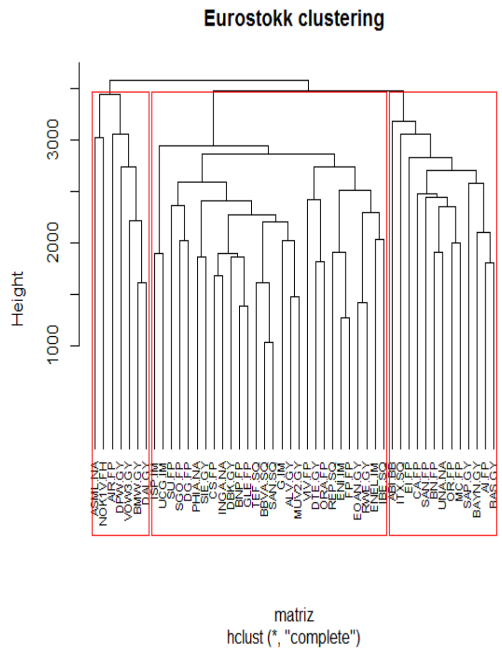


Figure 1 – MST 2002-2014 – Case1 *Figure 2 – HT 2002-2014 - Case 1*

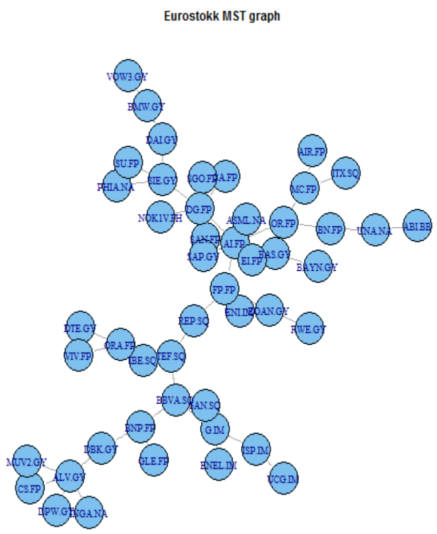
The MST allows revealing geometrical aspects of the asset returns and volume trading. The length of the link between connected firms is proportional to the distance between them and the geometrical aspect of the MST reveals the possible intermediate connections between any two firms of the group.

The hierarchical tree obtained starting from the MST described in Figure 1 is shown in Figure 2. In the figure, each vertical line indicates a firm. The height of the horizontal line indicates the ultrametric distance at which the two firms are connected. Each of the investigated stocks is indicated with its tick symbol in the figure.

In period 2002-2014, three clusters are clearly identified. The central cluster is composed by eleven French companies (AIR LIQUIDE, AXA, BNP PARIBAS, CARREFOUR, SAINT GOBAIN, L'OREAL, LVMH MOET HENNESSY, SCHNEIDER ELECTRIC, GRP SOCIETE GENERALE, VINCI and VIVENDI); five German companies (ALLIANZ, BASF, DEUTSCHE BANK, MUENCHENER RUECK and SIEMENS), three Spanish (BCO BILBAO VIZCAYA ARGENTARIA, BCO SANTANDER and TELEFONICA), two Dutch (ING GRP, PHILIPS) and one Italian company (ASSICURAZIONI GENERALI). Distribution by country of the second cluster is not so clear, being composed by ANHEUSER-BUSCH INBEV (BE), ASML HOLDING NV (NL), AIRBUS GROUP NV and ESSILOR INTERNATIONAL SA (FR), INDUSTRIA DE DISEÑO TEXTIL SA (ES) and NOKIA (FI). The last cluster is integrated mainly by German companies (BAYER, BMW, DAIMLER, DEUTSCHE POST, DEUTSCHE TELEKOM AG, EON AG, RWE, SAP and VOLKSWAGEN PREF), followed by the French (DANONE, ORANGE, SANOFI and TOTAL), and Italian companies (ENEL SpA, ENI SpA, INTESA SANPAOLO and UNICREDIT), two Spanish (IBERDROLA and REPSOL) and one Dutch company (UNILEVER NV).

Notwhistanding the segregation by country, we have to stand out that the business sector of the first cluster is mainly banking (six companies), insurance (four companies) and industrial (three companies) while the third cluster has a relevant composition of utilities companies (four) while oil & gas and automobiles & parts (three companies each).

Figure 3 shows the MST in the critical situation for the whole period (Case 5). Studying the critical situation is important in order to detect changes in the market structure. In particular, this could be interesting during crisis events, detecting different clusters configuration.



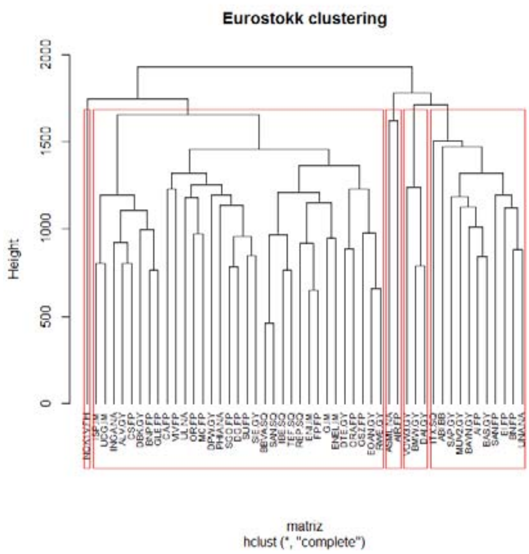


Figure 3 – MST 2002-2014 – Case 5

Figure 4 – HT 2002-2014 - Case 5

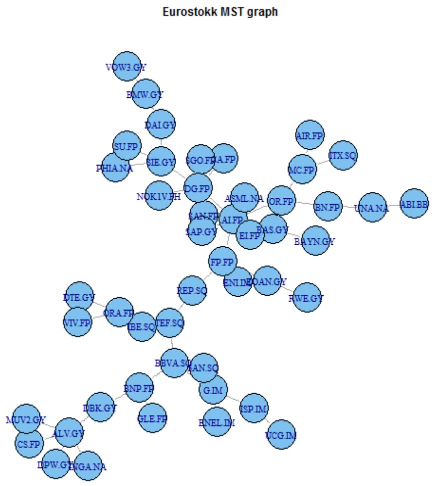
Note that moving from a *normal* to a *critical* situation generates a bigger number of clusters. In fact five clusters are determined, with a less than half diameter with respect to the normal situation. AIR LIQUIDE is the most connected stock, with sixteen links with other stocks while TELEFONICA is the second one, with five links. Figure 4 shows the HT in the critical situation.

After analyzing our complete time period, we now focus on the period from 2008 on. In the normal situation, the 2008-2014 sub-period (Case 3) has a greater number of clusters than in 2002-2007 sub-period (Case 2). Inversely, in critical codification, the number of clusters diminishes from 2002-2007 (Case 6) to 2008-2014 (Case 7), showing more co-movement between stocks. Particularly, in sub-period 2008-2014, this greater co-movement is showed with a star-shaped MST, in the *critical* codification case. In that sub-period, shares are mostly situated around AIR LIQUIDE[5].

While for normal encoding, the MST is more widespread, in the *critical* situation the MST shows more concentration, evidencing a greater bonding between stocks. Figure 5 and 6 show

[5] Case 4 and Case 8, resulting from considering two additional stocks with data belonging only to 2008-2014 sub-period, didn't show significant differences with Case 3 and Case 7, where 48 stocks were analyzed.

the MST and HT for the 2008-2014 sub-period, using the *normal* codification, with forty-eight variables.



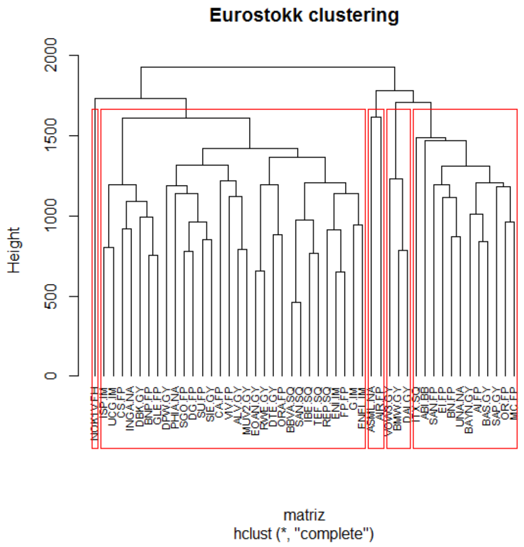
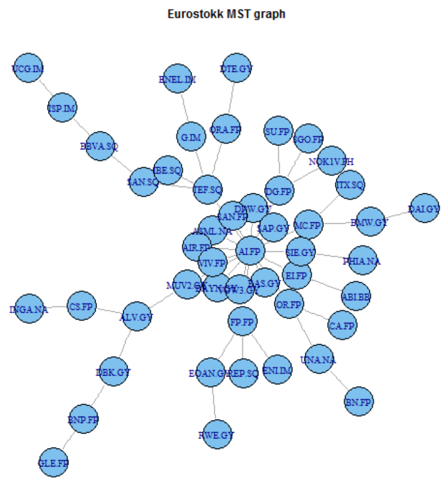


Figure 5 – MST 2008-2014 – Case 3

Figure 6 – HT 2008-2014 - Case 3

Figure 7 and 8 show the MST and HT for the 2008-2014 sub-period, applying the *critical* codification, with forty-eight variables.



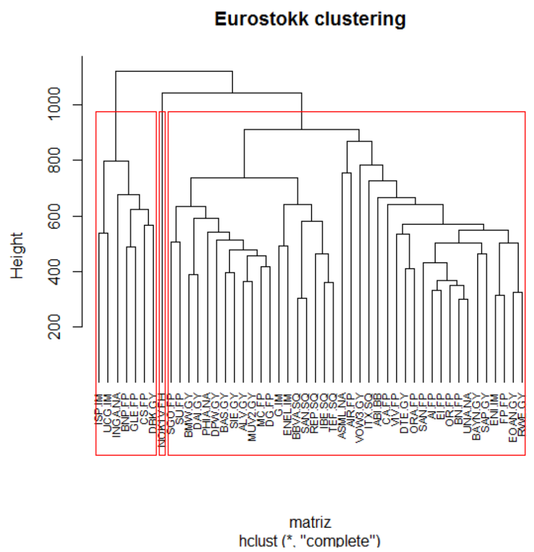


Figure 7 – MST 2008-2014 – Case 7

Figure 8 – HT 2008-2014 - Case 7

Figure 7 corresponds to Case 7 which, together with Case 8, showed AIR LIQUIDE as the most connected stock, with eighteen and seventeen links, respectively. In spite of AIR LIQUIDE not being the biggest or the best reputed within Eurostoxx companies, it has shown the lowest volatility in the whole period (see table 2). In this fashion, it seems that during crisis periods the stock network tends to center around the most stable asset.

On the other hand, although the number of clusters varied, several groups of stocks stayed together in the different sub-periods analyzed. Firstly, companies that compose these groups have in common that they belong to the same country: this is the case with large groups of French and German stocks and also of Spanish and Italian companies. Secondly, it is also relevant to note that several sectorial groups kept together in the whole period. The largest group is composed by banks (between six and eight companies), industrial companies (the four stocks included in the index), utilities companies (the four stocks analyzed), insurance companies (between three and four stocks), chemicals, telecommunications, oil & gas and personal & household goods (each with three companies).

In sum, regardless of the type of codification, there were several subsets of stocks that kept together in all periods analyzed. From this point of view, the grouping by country continues to be very relevant, as well as the industry grouping. Worth of highlighting is the grouping of companies from France, Germany, Italy and Spain, as well as groups of banks, utilities, industrial companies, to name just a few.

Two main results can be inferred from the previous analysis. First, when moving from the normal to the critical situation for all analyzed periods it can be observed that clusters tend to maintain their composition at least for some stocks. Concretely, there is an apparently "country-specific" configuration of clusters more distinguished than a sectorial configuration. This fact that both stock prices and volumes seem to be country specific aside from following a sectorial pattern as normally happen in single-country markets (see Mantegna (1999); Kocakaplan et al. (2013); Peron et al. (2012); Preis et al. (2012)) may suggest the presence of insufficient integration of the European stock markets. Additionally, this arrangement of stocks is more intense during 2008-2014. For the duration of this period, the most numerous cluster shows at least three sub-clusters (see Figure 8). One of this sub-clusters includes all Spanish companies except INDITEX (ITX), while the other two sub-clusters contain a German group and a French group firms. This configuration suggests that during periods of economic crisis and high market volatility, the network organization increases its degree of "national" stock co-movements instead of its degree of "European" market. A second result is obtained when comparing the whole period under analysis (2002-2014) with the final selected sub-period (2008-2014), which corresponds to a period of crisis in Europe. In line with other studies (Heiberger, 2014; Wilinska et al. 2014), during this crisis the network changes into a more centralized structure shaped like a star. This form is revealed when using the critical codification to focus on high volatility phases. This centralized structure might be more prone to spread turbulences through the whole network. However, the very center of the network is occupied by the lowest volatility stock in the whole period (AIR LIQUIDE) suggesting a sort of stability anchor during this period of economic turbulence and uncertainty.

We now turn to the analysis of diameter of MST and number of clusters, as they evolve in time. Table 3 shows the results obtained for each of the windows identified, both for a critical and normal codification.

Table 3 – Dynamic analysis

Critical			Normal		
Año	Diameter	Clusters	Año	Diameter	Clusters
2002	1235	3	2002	1413	5
2003	1001	2	2003	1628	3
2004	390	2	2004	1057	6
2005	213	2	2005	999	3
2006	522	3	2006	1057	2
2007	451	3	2007	1621	3
2008	926	5	2008	1587	6
2009	781	5	2009	1540	2
2010	879	3	2010	1259	6
2011	617	6	2011	1370	4
2012	513	9	2012	1463	8
2013	570	8	2013	1572	4
2014	485	3	2014	1413	5

Source: Own elaboration using R

In most cases, number of clusters is bigger in *normal* codification cases with respect to *critical* situation cases. On the other hand, as would have expected, the diameter for critical situation is in every situation lower than that for normal cases. With the exception of 2002 and 2003, overall diameter for *normal* situation doubles or triples the diameter for *critical* situation cases. Figure 9 shows the evolution of diameter in both codifications for each of the rolling windows defined. Interestingly enough, this figure shows both lines diverging over time. This result suggests an increasing difference in network cross-correlations between *normal* and *critical* market stress. Additionally, as we have previously mentioned, clustering behavior observes a significant change when the financial crisis period is separately analyzed, moving to a more country group organization within the network. Putting together both results, it is clearly suggested that the diversification effect that should protect a portfolio have diminished over the time sample we analyze and especially as the financial crisis evolved just when it would have been most needed. Preis et al. (2012) obtain the same conclusion by analyzing cross-correlations between 72 Dow Jones stocks.

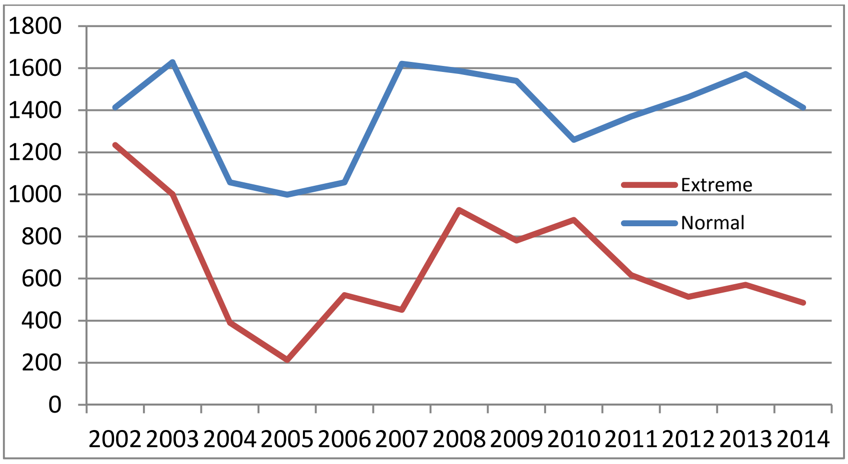


Figure 9 – Evolution of diameter for rolling windows

4. Conclusions

This paper follows the Symbolic Time Series approach for clustering introduced by Mantegna (1999), Brida and Risso (2007; 2009), Brida et al. (2009), Gullapalli and Carley (2013) and Abbasi and Loun (2014). We combine clustering and symbolizing methods to study topology and hierarchy in the Euro Stoxx 50 from 2002 up to 2014. This combination of methodologies has been followed by Brida´s research group and is itself a novelty in the more traditional approach pioneered by Mantegna (1999). Additionally, in this paper we introduce a multidimensional approach which permits clustering analysis putting together price returns and volumes in a single network. Finally, as long as we include in our time sample the period corresponding to the global financial crisis and subsequent European sovereign crisis, we explore on the topology the effects of this extreme economic events.

Three main results arise from this analysis. Firstly, it is apparent an increase in synchronization and connectivity in the Euro Stoxx 50 market during the global crisis, in line with similar analysis such as Preis et al. (2012) and Peron et al. (2012). Additionally, at this time, the network becomes a more centralized one, as other studies have highlighted (e.g., Heiberger, 2014; Wilinska et al., 2013). Secondly, during our second period of analysis, 2008-2014, we observe that hierarchy becomes more country-specific and different sub-clusters of stocks belonging to France, Germany, Spain and Italy are found apart from their activity sector. This result suggests that during this period of time financial investors seem to be most worried about country specific economic circumstances in line with complete different stories regarding the European sovereign crisis (Lane, 2012). However, we do not observe this result when the critical symbolization is computed (Figure 7 and 8). This two results are highlighting the fact that financial agents tend to pay attention to country specific risks during volatile times and financial crashes events and, therefore, they may be suggesting that European monetary and financial integration are far to be achieved.

Finally, we have shown that the diversification effect that should protect a portfolio have diminished over the time sample, especially from the outbreak of the financial crisis on. Differences between cross-correlations in normal and critical stages of market stress have increased over our time sample putting difficulties on the design of diversified and less risky portfolios. To the future, an innovative line of research could be to compare the results obtained using only prices, for which there are many papers, with those obtained using price and volume for the same dataset.

References

Abbasi, A. Q., & Loun, W. A. (2014). Symbolic Time Series Analysis of Temporal Gait Dynamics. *Journal of Signal Processing Systems* 74(3): 417-422.

Barone, E. (1990). The Italian Stock Market: Efficiency and Calendar Anomalies. *Journal of Banking and Finance* 14: 483-510.

Birch, J., Pantelous, A. A., & Soramäki, K. (2015). Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns. *Computational Economics* DOI 10.1007/s10614-015-9481-z.

Bonanno, G., Caldarelli, G., Lillo, F., Miccichè, S., Vandewalle, N., & Mantegna, R. N. (2004). Networks of equities in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems* 38(2): 363-371.

Bonanno, G., Caldarelli, G., Lillo, F., & Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E* 68(4): 1-4.

Bonanno, G., Lillo, F., and Mantegna, R.N., (2001). High-frequency cross-correlation in a set of stocks, *Quantitative Finance* 1: 96–104

Bouchaud, J. and Potters, M. (2000). *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge.

Brida, J. G., Matesanz, D. and Risso, W. (2009). Dynamical Hierarchical Tree in Currency Markets. *Expert Systems with Applications* 36: 7721-7728.

Brida, J. G., & Risso, W. A. (2009). Dynamic and Structure of the Italian Stock Market based on returns and volume trading. *Economics Bulletin* 29(3): 2420-2426.

Brida, J. G. and Risso, W. (2007). Dynamics and Structure of the Main Italian Companies", *International Journal of Modern Physics C* 18(11): 1783 – 1793.

Brida, J. G., Puchet Anyul, M. and Punzo, L.F. (2003). Coding economic dynamics to represent regime dynamics: a teach-yourself exercise. *Structural Change and Economic Dynamics* 14: 133-157.

Burt, R., Carlton, D., 1989. Another Look at the Network Boundaries of American Markets. *The American Journal of Sociology* 95(3): 723-753.

Chi, K. T., Liu, J., & Lau, F. C. (2010). A network perspective of the stock market. *Journal of Empirical Finance* 17(4): 659-667.

Clark, P. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41: 135-155.

Cornell, B. (1981). The Relationship between Volume and Price Variability in Future Markets. *The Journal of Future Markets* I: 303-316.

Crouch, R. (1970). The Volume of Transactions and Price Changes on the New York Stock Exchange. *Financial Analysis Journal* 26: 104-109.

Daw C.S., Finney, C.E.A., Tracy, E.R. (2003). A review of symbolic analysis of experimental data. *Review of Scientific Instruments* 74: 916-930.

De Polsi, G., Cabeza, C., Marti, A. C., & Masoller, C. (2013). Characterizing the dynamics of coupled pendulums via symbolic time series analysis. *The European Physical Journal Special Topic*, 222(2): 501-510.

Dias, J. (2012). Sovereign debt crisis in the European Union: A minimum spanning tree approach. *Physica A: Statistical Mechanics and its Applications* 391: 2046-2055.

Feser, E., Sweeny, S. (2000). A Test for the Coincident Economic and Spatial Clustering of Business Enterprises. *Journal of Geographical Systems* 2: 349-373.

Godfrey, M. D.; Granger, C. W. J. and Morgenstein, O. (1964). The Random Walk Hypothesis of Stock Market Behavior. *Kyklos* 17: 1-30

Granger, C. W. J., and Morgenstern, O. (1963), Spectral Analysis of New York Stock Market Prices. *Kyklos* 1: 1-27

Gullapalli, A., & Carley, K. M. (2013). Extracting ordinal temporal trail clusters in networks using symbolic time-series analysis. *Social Network Analysis and Mining* 3(4): 1179-1194.

Hakansson, H., Ford, D., (2002). How Should Companies Interact in Business Networks? *Journal of Business Research* 55: 133-139.

Halinen, A., Tornroos, J. (1998). The Role of Embeddedness in the Evolution of Business Networks. *Scandinavian Journal of Management* 14(3): 187-205.

Harris, L. (1986). Cross-Security Tets of the Mixture of Distribution Hypothesis. *Journal of Financial and Quantitative Analysis* 21: 39-46.

Harris, L. (1983). The Joint Distribution of Speculative Prices and of Daily Trading Volume. Working Paper No 38-84, Department of Finance and Business Economics, University of Southern California, Los Angeles, CA.

Henderson, J., Dicken, P., Hess, N., Yeung, H. (2002). Global Production Networks and the Analysis of Economic Development. *Review of International Political Economy* 9: 436-464.

Heiberger, R. H. (2014). Stock network stability in times of crisis. *Physica A: Statistical Mechanics and its Applications* 393: 376-381.

Holton, G., 2003. *Value at Risk: Theory and Practice*, ELSEVIER Academic Press.

Jain, P. and Joh, G. (1988). The Dependence between Hourly Prices and Trading Volume. *Journal of Financial and Quantitative Analysis* 23: 269-284.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Derivatives Risk*, McGraw-Hill

Karpoff, J. (1987). The relation between price change and trading volume: A survey. *The Journal of Financial and Quantitative Analysis* 22(1): 109–126.

Keskin, M., Deviren, B., & Kocakaplan, Y. (2011). Topology of the correlation networks among major currencies using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications* 390(4): 719-730.

Kocakaplan, Y., Doğan, Ş., Deviren, B., & Keskin, M. (2013). Correlations, hierarchies and networks of the world's automotive companies. *Physica A: Statistical Mechanics and its Application*s 392(12): 2736-2774.

Lane, P. R. (2012). The European Sovereign Debt Crisis. *Journal of Economic Perspectives* 26(3): 49–68

Lo, A. and Wang, J. (2000). Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. *The Review of Financial Studies* 13(2): 257-300.

Lyócsa, Štefan, Tomáš Výrost, and Eduard Baumöhl. (2012). Stock market networks: The dynamic conditional correlation approach. *Physica A: Statistical Mechanics and its Applications* 391(16): 4147-4158.

Makki Alamdari, M., Samali, B., & Li, J. (2015). Damage localization based on symbolic time series analysis. *Structural Control and Health Monitoring* 22(2): 374-393.

Mantegna, R. (1999). Hierarchical Structure in Financial Markets. *The European Physical Journal B* 11: 193-197.

Mantegna, R. N., and Stanley, H. E. (1999), *Introduction to Econophysics.* Cambridge University Press, Cambridge, UK.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7: 77-91.

Michaely, R. and Murgia, M. (1995). The Effect of tax heterogeneity on prices and volume around the Ex-dividend day: Evidence from the Milan Stock Exchange. *The Review of Financial Studies* 8(2): 369-399.

Molgedey, L. and Ebeling, W. (2000). Local Order, Entropy and Predictability of Financial Time Series. *The European Physical Journal B* 15: 733-737.

Morgan, I. (1976). Stock Prices and Heteroskedasticity. *Journal of Business* 49: 496-508.

Oatley, T., Winecoff, W., Pennock, A., & Danzman, S. B. (2013). The political economy of global finance: a network model. *Perspectives on Politics* 11(01): 133-153.

Onnela, J.P., Chakraborti, A., and Kaski, K. (2003) Dynamics of market correlations: taxonomy and portfolio analysis. *Physical Review E* 68: 056110.

Onnela, J.P., Kaski, K., and Kertesz, J. (2004). Clustering and information in correlation based financial networks. *The European Physical Journal B-Condensed Matter and Complex Systems* 38(2): 353–362.

Osborne, M. (1959). Brownian motion in the Stock Market. *Operations Research* 7: 145-173.

Pérez, J. (2005). Empirical identification of currency crises: differences and similarities between indicators. *Applied Financial Economics Letters* 1: 41-46.

Podobnika, B., Horvati, D., Petersen, A. M. and Stanley, H. E. (2009). Cross-correlations between volume change and price change, *Proceedings of the National Academy of Sciences* 106(52): 22079–22084.

Peron, T.K., Costa, L.F., and Rodrigues, F.A. (2012). The structure and resilience of financial market networks. *Chaos* 22: 013117.

Preis, T., Kenett, D.Y., Stanley, H. E., Helbing, D., and Ben-Jacob, E. (2012). Quantifying the Behavior of Stock Correlations under Market Stress. *Scientific Reports* 2: 752.

Ramal, R., Toulouse, G. and Virasoro, M. (1986). Ultrametricity for physicists. *Review of Modern Physics*, 58(3): 765–788.

Rogalski, R. (1978). The Dependence of Price and Volume. *The Review of Economics and Statistics* 36: 268-274.

Sorrentino, T., Aragoneses, A., Perrone, S., Gauthier, D. J., Torrent, M. C., & Masoller, C. (2014). Experimental study of the complex dynamics of semiconductor lasers with feedback via symbolic time-series analysis. In *SPIE Photonics Europ*e (pp. 91340L-91340L). International Society for Optics and Photonics.

Souma, W., Aoyama, H., Fujiwara, Y., Ikeda, Y., Iyetomi, H., Kaizoji, T. (2006). Correlation in Business Networks. *Physica A: Statistical Mechanics and its Applications* 370: 151-155.

Souminen, M. (2001). Trading Volume and Information Revelation in Stock Markets. *The Journal of Financial and Quantitative Analysis* 36(4): 545-565.

Tauchen, G. and Pitts, M. (1983). The Price Variability-Volume Relationship on Speculative Markets. *Econometrica* 51: 485-505

Tumminello, M., Aste, T., Di Matteo, T., and Mantegna, R.N., (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences* 102: 10421–10426.

Tse, C. K., Liu, J., & Lau, F. C. (2008). Winner-take-all correlation-based complex networks for modeling stock market. *Proceedings of the International Symposium on Nonlinear Theory and its Applications*. Budapest, Hungary.

Vandewalle, N., Brisbois, F., & Tordoir, X. (2001). Non-random topology of stock markets. *Quantitative Finance* 1(3): 372-372.

Westerfield, R. (1977). The Distribution of Common Stock Price Changes: An Application of Transactions Time and Subordinated Stochastic Models. *Journal of Financial and Quantitative Analysis* 12: 743-765.

Wilińskia, M., Sienkiewicza, A., Gubieca, T., Kutnera, R., and Struzikb, Z.R. (2013). Structural and topological phase transitions on the German Stock Exchange. *Physica A: Statistical Mechanics and its Applications* 392: 5963–5973.