CNMV
Comisión Nacional del Mercado de Valores

Network-based Measures as Leading Indicators of Market Instability: The case of the Spanish Stock Market

Gustavo Peralta

Documentos de Trabajo
Nº 59

Network-based Measures as Leading Indicators of Market Instability: The case of the Spanish Stock Market*

Gustavo Peralta†

Documentos de Trabajo

Nº 59

Marzo/March 2015

* I thank Pablo Gasós y Casao and Ricardo Crisóstomo Ayala for their helpful comments.

† Gustavo Peralta (corresponding author) is at the Department of Research and Statistics at CNMV and Department of Business Administration at Universidad Carlos III de Madrid. Email: gperalta@cnmv.es. The opinions in this Working Paper are the sole responsibility of the author and they do not necessarily coincide with those of the CNMV.

Gustavo Peralta is a member of the Research, Statistics and Publications Department, CNMV.

The opinions in this Working Paper are the sole responsibility of the authors and they do not necessarily coincide with those of the CNMV.

The CNMV publishes this Working Paper Series to enhance research and contribute towards greater knowledge of the stock markets and their regulation.

The CNMV distributes its reports and publications via the Internet at www.cnmv.es

© CNMV. The contents of this publication may be reproduced, subject to attribution.

Gustavo Peralta pertenece al Departamento de Estudios, Estadísticas y Publicaciones de la CNMV.

Las opiniones expresadas en este documento relejan exclusivamente el criterio de los autores y no deben ser atribuidas a la Comisión Nacional del Mercado de Valores.

La Comisión Nacional del Mercado de Valores, al publicar esta serie, pretende facilitar la difusión de estudios que contribuyan al mejor conocimiento de los mercados de valores y su regulación.

La Comisión Nacional del Mercado de Valores difunde la mayoría de sus publicaciones a través de la red Internet en la dirección www.cnmv.es

© CNMV. Se autoriza la reproducción de los contenidos de esta publicación siempre que se mencione su procedencia.

ISSN (edición electrónica): 2172-7147

Maqueta: Composiciones Rali, S.A.

Abstract

This paper studies the undirected partial-correlation stock network for the Spanish market that considers the constituents of IBEX-35 as nodes and their partial correlations of returns as links. I propose a novel methodology that combines a recently developed variable selection method, Graphical Lasso, with Monte Carlo simulations as fundamental ingredients for the estimation recipe. Three major results come from this study. First, in topological terms, the network shows features that are not consistent with random arrangements and it also presents a high level of stability over time. International comparison between major European stock markets extends that conclusion beyond the Spanish context. Second, the systemic importance of the banking sector, relative to the other sectors in the economy, is quantitatively uncovered by means of its network centrality. Particularly interesting is the case of the two major banks that occupy the places of the most systemic players. Finally, the empirical evidence indicates that some network-based measures are leading indicators of distress for the Spanish stock market.

Keywords: Network Theory, Stock Markets, Systemic Risk Indicators.

JEL Classification: G01, G12, G17, C45, C58.

Table of contents

Index of figures

Index of tables

1 Introduction

Nobody can disagree that we live a networked life. Our interactions are embedded in different systems that can be thought as networks. Examples include the social network of real and virtual acquaintance (Twitter, Facebook), the Internet, the World Wide Web, distribution networks such as airport transportation flow or postal delivery routes, the economic and financial network, the network of academic citation and the list is easily extendable, Newman (2003). As a consequence of the recent financial crisis, the fields of economics and finance have captured the attention of network researchers aiming to revise and extend established theoretical frameworks. Schweitzer et al. (2009) presents the future challenges for this discipline while Allen and Babus (2009) and Jackson (2014) set out the salient theoretical findings achieved so far. Additionally, prominent policymakers, such as the Executive Director of the Financial Stability Department at the Bank of England (at the time the speech was made) Haldane (2009) and the current Chair of the Board of Governors of the Federal Reserve System, Yellen (2013), have argued in favor of the network approach as a valuable tool to enhance our understanding of excessive systemic risk in the financial system.

Recently, several papers studying stock markets through the lens of network theory have been published in prestigious academic journals. The current paper contributes to this branch of literature in three directions. First, I propose an enhanced and convenient methodology for estimating a network based on partial correlation for stocks that are listed in a particular market. This methodology is grounded in a high-dimensional setting and allows researchers to control for the statistical significance of the underlying network. Second, since the mainstream literature is centered in the US market, the current paper contributes to this line of research by targeting the Spanish stock market. The aim is to develop new insights into its topological structure and to quantitatively identify its systemic players. Additionally, four other European stock markets are included in the study which allows us to compare their salient characteristics under the same unified framework. Finally, I provide an empirical assessment accounting for the role of some network-based measures as leading indicators of financial distress for the Spanish market.

Specifically, I estimate the Spanish *Partial-Correlated Stock Network (PCSN)* in which each node corresponds to a stock comprising the market index, IBEX-35, and the links between them account for their return's partial correlation. I use a daily dataset comprising a 10 year sample period starting in Nov-2004 until Sep-2014, thus covering both tranquil and crisis periods. Since a sparse partial correlation matrix is required in order to detect the skeleton of the market, I estimate such a matrix by applying a recently developed *Graphical Lasso* algorithm by Friedman, Hastie, and Tibshirani (2008). Methodologically, the paper presents some improvements when it is compared to the mainstream literature solving, or at least attenuating, three of its drawbacks. First, partial correlations of returns are computed instead of direct correlations. This enables us to calculate the co-movements between two stocks alone while controlling for the effects exerted by the rest of the firms in market. In this regard, a connection between a pair of assets exists as long as they are exposed to common factors different from the general market factor. Second, Graphical Lasso is a totally automatic and data-driven technique which is grounded in a high-multivariate setting. Such a technique permits the estimation of *sparse* partial correlation matrices which in turns allows us to uncover the underlying network structure. Since there is no need to use ad-hoc topological constraints or arbitrarily predefined thresholds, as it is the case with traditional methodologies, the resulting structure is not distorted by artificial restrictions. Finally, Monte Carlo simulations are implemented in order to determine the statistical significance of the related network.

Three empirical results come from this study. First, the main indicators of the Spanish *PCSN* show features that are inconsistent with networks made by chance. In particular, the evidence supports a fairly stable structure with high levels of transitivity and smaller connected components.[1] In addition, four large European markets, namely the UK, France, Germany and Italy, are included in the analysis for comparison purposes. Following the exact same approach for all of them, the evidence shows an astonishing similarity in terms of their topological organization.

Second, an in-depth study of the systemic importance of different economic sectors is presented for the abovementioned stock markets. The quantification of the systemic importance of sectors and stocks is based on the notion of network centrality, a concept intensively used in the sociological literature. In general terms, the data shows a positive and statistically significant relationship between market capitalization and stock centrality after controlling for the effects of economic sectors. Moreover, the importance of economic sectors varies across stock markets since the *PCSN* of Spain, France and Italy are characterized by a greater centrality of the banking industry. The cases of Germany and UK are different since for those structures it is the utility and industrial sector, respectively, the most influential ones. For specific case of the Spanish network, it is particularly interesting to mention the disproportionally large centrality scores shown by two financial firms, Banco Santander and BBVA. This observation is consistent with the conventional wisdom and allows us to consider them as the most systemic players in the Spanish market in a very specific and quantitative fashion.

Third, I investigate the extent to which network-based measures are leading indicators of market distress by means of two complementary approaches. As a first approach I estimate a Probit model where the dependent variable assumes value one when the IBEX-35 drops by more than 3 standard deviations and zero otherwise. The set of independent variables includes lagged values of some network measures; say d*ensity, transitivity* and the *centrality of the banking sector*. The result shows that the probability that the Spanish market suffers from large negative movement increases when the lagged network becomes denser or when the banking industry scores high in centrality. In the second approach it is assumed that the return pro-

1 A list with definitions of network-related concepts is provided in the body of the paper.

cess of the market index follows an ARCH model. In this model the specification of the conditional market volatility includes as independent variables the same set of lagged network measures as in the previous approach. The estimation shows that the market variance increases with network density and with the centrality of the banking sector while it reduces with the level of transitivity.

The remainder of the paper is organized as follows. Section 2 presents a literature review regarding the current study. Section 3 defines the Partial-Correlated Stock Network and some network-based measures that are going to be used through the paper. Section 4 describes the dataset used and the estimation methodology. Section 5 establishes the main results from the estimation procedure distinguishing between a static and dynamic analysis. Section 6 provides the statistical assessment of some network-based measures acting as leading indicators of market distress. Finally, section 7 concludes and outlines future research lines.

2 Literature Review

The literature regarding the study of stock markets throughout network theory can be divided into three sub-branches according to their main focus. The first group includes those studies focused purely on the topological description of the related network and is basically grounded on physics literature. A second group of articles, closely linked to the econometric and financial research, relies on the network approach to get new insights about systemic risk issues. Finally, recent developments have used network theory as a promising tool to enhance portfolios' performances. Next, I briefly review the major features from each of these research lines.

Physics literature has been shown to be very productive in using network concepts to describe stock markets. Mantegna (1999) and Bonanno, Lillo, and Mantegna (2001) were among the first in this endeavor, applying a particular correlation-based filtering procedure, the so-called *Minimum Spanning Tree (MST)*, for the US market in order to study its skeleton.[2] The authors find a hierarchical structure in the stock network where its branches were associated with specific economic sectors. In similar vein, Vandewalle, Brisbois, and Tordoir (2001) estimate the *MST* also for the US stock market providing evidence for power-law degree distributions and connectivity patterns that are inconsistent with random networks. For applications of this approach to markets other than the US or with broader datasets, see Jung et al. (2006), Garas and Argyrakis (2007) and Huang, Zhuang, and Yao (2009).

Two variations of this baseline framework are noted. On the one hand, Onnela et al. (2003) studies the time-dependent properties of the *MST* for the US market implementing a moving window approach. The authors call this methodology *Dynamic Asset Trees*. On the other hand Onnela et al. (2003) proposes a *threshold approach* where correlations below a pre-established and arbitrarily chosen threshold are discarded for the construction of the network. A recent empirical application of this methodology to the US market is provided in Tse, Liu, and Lau (2010) where power law degree distributions are also reported for sufficiently large value of the threshold. For authoritative summaries of this research line see Bonanno et al. (2004) and Tumminello, Lillo, and Mantegna (2010).

The last years have witnessed increased interest in network theory among the financial research community as a way to shed some light on systemic risk issues. In Billio et al. (2012) the authors build a directed Granger-causality network.[3] Such a

2 MST is a filtering technique that allows us to build a connected network of N stocks by joining together pairs of them according to their pair correlation (in decreasing order) as long as no loops are formed in the structure. The resulting network is a tree network.

3 The authors also measure connectedness through principal components but this approach is not the focus of the current paper.

structure captures the market interconnectedness between financial institutions where links account for statistically significant lead-lag relationships between monthly returns. Among their results they show *i)* how the system became much more interconnected during and before the last financial crisis *ii)* evidence on how central institutions were the ones with the largest financial losses. Following a similar approach with strong emphasis on connectedness, Diebold and Yılmaz (2014) proposes a network in which connections between financial institutions are assigned according to the variance decomposition of the volatility forecast error. This approach gives rise to a volatility weighted-directed network finding a clustered interaction between government-sponsored firms and investment banks. Additionally, the authors also demonstrate how the cycles of the total connectivity in such structure coincide with major disruptions in the US market.

Some limitations regarding the study of Billio et al. (2012) are noted. On the one hand, pairwise Granger tests could lead to misleading results in a multivariate context. On the other hand, not accounting for correlation in the tails of the return's distributions undermines systemic risk conclusions. These concerns are taken into account in Hautsch, Schaumburg, and Schienle (2014) by estimating a *tail-risk network*, a weighted-directed network in which the links between institutions are given by the interconnectedness of firms' Value-at-Risk. With this network in place, the authors compute the systemic relevance of financial firms in terms of their destabilizing power, the so-called *realized systemic risk beta*. In a closely related paper Hautsch, Schaumburg, and Schienle (2014), the authors adapt the *tail-risk network* framework to forecast firms' systemic relevance.

A far less explored research area relates to the use of network theory as a way to support investment choices. As far as I am aware, there are just two articles following this line of research, and both use *MST* in combination with shrinkage covariance estimation techniques as methodological tools. Aiming to build well-diversified portfolios, Pozzi, Di Matteo, and Aste (2013) show the improvement in financial performance of an investment strategy that assigns wealth toward stocks belonging to the periphery of the stock network, the poorly connected stocks. The main assumption in Pozzi, Di Matteo, and Aste (2013) is that the individual dimension of stocks (e.g. Sharpe ratio) is uncorrelated with their systemic dimension (its centrality score in the network). In Peralta and Zareei (2014) the authors show that this correlation is time and market dependent. Therefore, there are times in which central stocks also present good individual performances giving rise to a trade-off in the portfolio selection process. This fact leads them to create the so-called *ρ-dependent strategy* and to obtain enhanced out-of-sample results out of its implementation compared to well-known benchmarks. Finally, it worth mentioning the salient results in Ozsoylev, Walden, and Yavuz (2014). They study informational diffusion in the Istanbul Stock Exchange by using a network where investors (not financial firms) are the nodes whereas links relate to the channels through which information flows. Among their major contributions it should be remarked that the central investors receive information signals earlier than peripheral ones allowing them to benefit from early trading advantages and higher returns.

3 Networks in the Context of Stock Markets

3.1 Defining the Partial-Correlated Stock Network

In general terms, a network is a pair of sets $\varphi = \{N, a\}$, with $N = \{1, 2, \ldots, n\}$ the set of nodes and a the set of links connecting pair of them. Then, if there is a link from node i to node j, $(i,j) \in a$. A convenient way to arrange the information contained in a is by means of the so-called adjacency matrix $A = [A_{ij}]$. A is a $n \times n$ matrix in which $A_{ij} \neq 0$ captures the existence of a relationship between node i and node j. The network is said to be *undirected* if $A = A^T$, therefore if $(i,j) \in a$ also implies $(j,i) \in a$. Note that for undirected network, no causal relationship is attached to links and they are visually represented as a line, $(j - i)$. On the other hand, if $A \neq A^T$, the network is said to be *directed* and A_{ij} entails a causal relationship from node j to node i which does not necessarily imply the reverse. In this case, the links are visually represented as arrows, $(j \rightarrow i)$. Further, if $A_{ij} \in \{0,1\}$, φ is said to be *unweighted*. However, when $A_{ij} \in \mathbb{R}$, such a link also carries information about the intensity in the interaction between nodes leading to a *weighted* network. The reader is referred to Newman (2010) and Jackson (2010) for a comprehensive treatment of the field.

Before discussing the construction of the Partial-Correlated Stock Network (*PCSN*), it is convenient to refresh the concept of partial correlation between pairs of random variables. Let us assume $\boldsymbol{r} = (r_1, \ldots, r_n)^T$ to be a random vector following a multivariate normal distribution with mean vector μ and covariance matrix Σ. The *partial correlation* between r_i and r_j, denoted by ρ_{ij}, quantifies the correlation between these two variables conditional on the rest. In this Gaussian environment, it is well known that $\rho_{ij} = 0$ implies conditional independency between i and j. The inverse of the covariance matrix (commonly named as the precision matrix), denoted by $\Omega = \Sigma^{-1} = [\omega_{i,j}]$, contains the fundamental information regarding the partial correlations matrix ρ as follows.

$$\rho = [\rho_{ij}] = -\Delta\Omega\Delta \tag{1}$$

where $[\Delta]_{ij} = 1/\sqrt{\omega_{ii}}\ for\ i = j\ and [\Delta]_{ij} = 0\ for\ i \neq j$. Simple calculations show that the elements of the main diagonal of ρ are (-1) which is meaningless in our framework and therefore I set them equal to zero. It is also important to mention the close connection between ρ and regression analysis as it is stated in expression (2) below (Stevens (1998)). In such equation β_{ij} corresponds to the coefficient in a regression where r_i is the dependent variable and $r_{-(i)} = \{r_k : 1 \leq k \leq n\ and\ k \neq i\}$ are the independent ones. Therefore, a zero regression coefficient indicates zero partial correlation and thus conditional independency. In expression (2) σ_{ei} accounts for the standard deviation of the residual in the regression where r_i is the dependent variable (the reader is referred to Appendix A for the formal proof).

$$\rho = \begin{bmatrix} -1 & \beta_{12}\frac{\sigma_{e2}}{\sigma_{e1}} & \cdots & \beta_{1n}\frac{\sigma_{en}}{\sigma_{e1}} \\ \beta_{21}\frac{\sigma_{e1}}{\sigma_{e2}} & -1 & & \beta_{2n}\frac{\sigma_{en}}{\sigma_{e2}} \\ \vdots & & \ddots & \vdots \\ \beta_{n1}\frac{\sigma_{e1}}{\sigma_{en}} & \beta_{n2}\frac{\sigma_{e2}}{\sigma_{en}} & \cdots & -1 \end{bmatrix} \tag{2}$$

Grounded on the Gaussian Graphical Models literature Whittaker (1990), I define two types of undirected *PCSN* depending on whether the focus is on the weighted or on the unweighted version. The weighted PCSN, $\varphi^w = \{N, \rho\}$, with N as the set of stock under study and the partial correlation matrix of stock's returns taking the place of its adjacency matrix.[4] Therefore, for $\rho_{ij} \neq 0$, there is a link connecting stock i and j with intensity ρ_{ij}. The need to include zeros outside the main diagonal of ρ (sparsity) is evident since a fully connected network comprises too much information to be analyzed. Then, sparsity in the partial correlation matrix is an essential feature to be investigated for construction of a valuable and informative *PCSN*, an aspect that is considered in the estimation methodology. The unweighted *PCSN*, $\varphi^u = \{N, \varrho\}$ consists of the same set of stocks while its corresponding adjacency matrix $\varrho = [\varrho_{ij}]$ is given by the following rules:

$$\varrho_{ij} = \begin{cases} 1, & \rho_{ij} \neq 0 \\ 0, & \rho_{ij} = 0 \end{cases} \tag{3}$$

3.2 Network-based Measures

It has been empirically proved that different sort of networks, ranging from biological to engineering networks, show astonishing similarity when some of their traits are compared. The cases of small world property, fat-tail degree distribution and high clustering are particularly popular in the network literature. Below I comment about those properties while providing a list of fundamental network measures that are going to be used through the paper. This list is by no mean complete and it should be considered as an attempt to enumerate the salient quantities describing any type of network.

Among the most basic network's concepts, *node-size* and *link-size*, called n and m respectively, account for the number of nodes and the number of links in the network. The *density* d measures the fraction of links that actually exist relative to the maximum possible links in the structure. In mathematical terms, $d = m / \binom{n}{2}$.

The *degree* of node i, k_i, is defined as the number of links attached to that node and the *mean degree* of the network, c, captures the average number per node in the structure. Formally, $c = 2m / N$. A closely related and fundamental concept describ-

4 As it was commented, for both φ^w and φ^u I discard the information in the main diagonal of the corresponding adjacency matrix to prevent uninformative self-loops.

ing any network is its *degree distribution*, *P(k)*, representing the empirical distribution of node degrees. That is, *P(k)* is the fraction of nodes in the network showing degree *k*. When *P(k)* is bell-shaped most of the nodes in that network show approximately the same degree or connectivity. However it is surprisingly common to observe fat-tailed degree distributions in real-world networks, Barabási and Albert (1999). This evidence highlights the increased probability of observing nodes that are relatively poorly connected co-existing with extremely well connected nodes, the hubs. It is typically to model such fat-tail degree distribution by assuming a power law form $P(k) = \gamma_0 k^{-\gamma_1}$.

A *path* between nodes i and j is a sequence of successive links $(i_1 i_2),(i_2 i_3),\ldots,(i_{t-1} i_t)$ such that each $(i_s i_{s+1}) \in a$ for $s \in \{1,\ldots,t-1\}$ with $i_1 = i$ and $i_t = j$. The length of such a path is the number of links traversed along that path. The *geodesic path* between nodes i and j is the shortest path between those nodes. The *diameter* of the network is the longest geodesic distance between any two nodes and the *mean distance* is the average over geodesic paths (note that the average distance is bounded above by the diameter). An interesting regularity observed in real-world networks is the so-called Small Worlds property. This property embodies the idea that the average distance and diameter is surprisingly small in comparison to its node's size. Technically, it is said that the mean distance scales logarithmically (or slower) with the node-size. For example, in the social network literature, the seminal paper Milgram (1967) give rise to the idea of 6 degrees of separation among any two persons in the world. This result also remains valid for Facebook since 5 degrees of separation was found in Ugander et al. (2011).

A *network* is said to be *connected* if there is a path connecting any two nodes in the structure, otherwise it is *disconnected*. When the network is disconnected, each subset of nodes forming a connected sub-network is called a *component*. In other words, a component is a subset of nodes where for each pair of nodes in that subset there exists at least one path connecting them. The typical network configuration corresponds to several components with just one of them fulfilling a large proportion of the structure. This large component is called the *largest component* and its size relative to the total number of nodes in the network is denoted by *L*. For Facebook, *L* attains a value of 99.91% which basically means that almost any two persons in this virtual world are reachable by following the correct path, Ugander et al. (2011).

Another measure worth defining is the *degree of assortativity Q*. If the correlation between the degrees of connected nodes is positive, high (low)-degree nodes tends to be connected with other high (low)-degree nodes. This tendency is called positive assortativeness or just *assortative* for short. An assortative network is expected to be arranged as a *core/periphery* structure where the core is composed by highly connected nodes and the periphery by poorly connected ones surrounding the core. For the case in which high-degree nodes tends to be connected with low-degree ones, the correlation between the degrees of connected nodes is negative and we called this tendency as negative assortativeness or *disassortative*. In this case, the general configuration of the network presents star-like features.

In mathematics, a relationship is said to be transitive when $A \to B$ and $B \to C$, then $A \to C$. In a network context, this relationship means that if node i is connected to j and j is connected to k, then i is connected to k. The level of network transitivity T

captures the likelihood that any given pair of nodes shares another common neighbor. Mathematically, T is calculated as the ratio between the number of triangles in the network divided by the total number of connected triples of nodes. In social networks, such a measure tends to be quite high since the probability that a pair of friends shows another friend in common is usually large. Particular interesting to mention is the fact that some researchers associate large values of T with a network structured in communities, Newman and Park (2003).

Most of the measures discussed so far are predominantly macroscopic in nature. However, there exist several measures capturing node's position in a given network. These sorts of measures, commonly referred to as centrality measures, are grounded in the sociological literature and their main goal is to account for the power or influence of each node in a given structure, Freeman (1978). Given the extension of this literature, I just comment about two of them that have become the standard in network analysis. The simplest node's centrality measure is *degree centrality* which accounts for the degree of a given node. Therefore, the importance of a node in a network is given by the number of links attached to it. This is a local measure that discards the information about the network structure beyond the "first friends" of a give node. A natural extension to the degree centrality is given by the *eigenvector centrality* due to Bonacich (1972) and Bonacich (2007). Its version for a weighted network case is in Newman (2004). In formal terms, the eigenvector centrality of node i, v_i, is proportional the weighted sum of the centralities of its neighboring nodes. Assuming the adjacency matrix of the network is ρ, the eigenvector centrality of node i is given by

$$v_i = \lambda^{-1} \sum_j \rho_{ij} v_j \tag{4}$$

Note that equation (4) shows that a highly central node becomes central either by being connected with many other nodes (degree centrality) or by being connected with highly central ones. By restating equation (4) in matrix terms, it can be seen that the vector of centrality scores, v, is given by the eigenvector of the adjacency matrix corresponding to the eigenvalue λ, where the largest eigenvalue is the preferred choice.[5]

$$\lambda v = \rho v \tag{5}$$

5 It is usual to report the vector of centrality normalized to one, this is also the case for the current paper.

4 Database Description and Estimation Methodology

The period of study covers almost 10 years of daily data starting from Nov 1, 2004 until Sep 30, 2014. I rely on Datastream for the information about dividends-and-split adjusted stock prices and returns. Additionally, market value and industrial sectors for the selected companies were also used throughout the analysis. The set of firms included into the sample comprises the constituents of the index IBEX-35 at September 2014. Table 1 reports the sample of firms including the corresponding ticker, their market value for the last day in the sample, the inception date in the index and descriptive statistics for their return distribution.

As it was commented previously, with the aim of capturing the backbone of the Spanish stock market, a sparse estimation of the partial correlation matrix is required. A dense partial correlation matrix would lead to a fully connected *PCSN* (a network in which each pair of nodes is connected) thus obscuring its salient features. The two most popular filtering procedures in the network literature are the *Minimum Spanning Tree (MST)* and the *Threshold Method (TM)*. Both of them present severe drawbacks. *MST* retains the highest correlations in accordance with a strong artificial topological restriction. In particular, this method only considers the largest correlations as links between stocks as long as the network becomes connected and no loops of order three are formed. Then, if a large association between two nodes exists but taking it into consideration implies a triangle in the network, it is discarded. The *TM* just consists of discarding as links in the network the correlations below a pre-specified threshold and thus, a disconnected structure could result. However, the determination of such a threshold represents its major weakness since it is usually pre-specified without any theoretical or statistical support. A final comment regards to the application of *MST* and *TM* to the direct correlation matrix of return instead of the partial correlation matrix as it is usually done in the literature. This is another inconvenience since a direct correlation between two stocks could be due to the effect of a third one affecting both simultaneously, and thus distorting the resulting pattern of interconnections.

In order to overcome these drawbacks and with the goal of obtaining a statistically validated sparse partial correlation matrix that translates into a clearer *PCSN*, a 2-Step procedure is proposed. In the first step, I rely on the Graphical Lasso algorithm developed by Friedman, Hastie, and Tibshirani (2008). This approach maximizes the penalized log-likelihood of a multivariate normal distribution with respect to the precision matrix, Ω.

$$\Omega_{G-Lasso} = \arg\max log\left|\Omega\right| - tr\left(S\Omega\right) - \gamma\left|\Omega\right|_1 \qquad (6)$$

where S is the sample covariance matrix, tr is the trace operator and $|\Omega|_1$ is the L_1-norm - the sum of the absolute values of the elements of Ω. The penalty or regularization parameter γ controls for the amount of shrinkage in the values of the components of Ω. When $\gamma = 0$, the estimation of Ω is just its maximum likelihood estimator, S^{-1}. When the amount of regularization increases, $\gamma > 0$, more parameters are pushed toward zero resulting in sparser solutions. Due to the particular form of the restriction in (6), exact zeros are found in the results of the algorithm. It is important to determine the value of the shrinking parameter in an optimal way. A tradeoff in determining γ exists since lower values of γ fit better to the data, at the cost of denser Ω. In the current setting γ is determined by 10-fold cross validation as it is proposed in Friedman, Hastie, and Tibshirani (2008).[6]

Descriptive statistics TABLE 1

Industry / Name	Ticker	Inception Date	Market Value* $	Market Value* %	Mean ** (1)	Median **	Std ** (2)	(1)/(2)	Min	Max
Industrial										
ABENGOA B SHARES	ABS	25/10/12	3,159	1%	33.2%	0.0%	46.3%	71.7%	-2,610.1%	2,952.2%
ABERTIS INFRAESTRUCTURAS	ACE		14,049	2%	8.4%	0.0%	24.8%	33.9%	-2,584.9%	3,040.8%
ACCIONA	ANA		3,393	1%	7.5%	0.0%	37.0%	20.3%	-3,262.9%	4,179.6%
ACS ACTIV.CONSTR.Y SERV.	ACS		9,690	2%	12.4%	18.7%	28.7%	43.1%	-2,523.3%	4,423.2%
AMADEUS IT HOLDING	AMS	29/04/10	13,257	2%	22.5%	18.5%	23.5%	96.1%	-1,791.9%	1,976.7%
ARCELORMITTAL (MAD)	MITT	28/07/06	18,153	3%	1.4%	0.0%	46.8%	3.0%	-4,830.1%	4,676.4%
DIST. INTNAC.DE ALIM	DIA	05/07/11	3,701	1%	21.1%	0.0%	28.8%	73.2%	-1,577.7%	3,076.9%
FERROVIAL	FERC		11,354	2%	11.8%	0.0%	35.0%	33.7%	-2,834.0%	3,502.2%
FOMENTO CONSTR.Y CNTR.	FCC		1,940	0%	-0.1%	0.0%	36.0%	-0.2%	-2,317.1%	3,639.3%
GAMESA CORPN.TEGC.	GAM		2,437	0%	8.6%	0.0%	45.1%	19.1%	-5,573.9%	5,519.3%
GRIFOLS ORD CL A	PROB	17/05/06	6,912	1%	27.1%	0.0%	30.7%	88.4%	-3,491.5%	2,992.2%
INDITEX	IND		68,177	12%	20.2%	0.0%	27.5%	73.5%	-2,580.6%	2,892.9%
INDRA SISTEMAS	IDR		1,822	0%	3.2%	0.0%	26.9%	11.8%	-1,903.3%	2,296.7%
JAZZTEL	JAZ		3,287	1%	22.1%	0.0%	49.4%	44.7%	-4,166.5%	6,666.6%
MEDIASET	TL5		4,009	1%	3.8%	0.0%	37.2%	10.3%	-3,426.3%	3,086.1%
OBRASCON HUARTE LAIN	OHL		2,642	0%	21.8%	0.0%	36.9%	59.1%	-2,817.7%	3,877.0%
REPSOL YPF	REP		25,385	4%	6.9%	0.0%	29.7%	23.1%	-3,938.5%	2,984.2%
SACYR	SCYR		2,141	0%	4.6%	0.0%	49.1%	9.4%	-3,111.4%	5,337.4%
TECNICAS REUNIDAS	TECN	21/06/06	2,347	0%	16.8%	5.2%	36.1%	46.6%	-3,260.9%	3,505.2%
VISCOFAN	VIS		2,023	0%	20.2%	0.0%	25.4%	79.7%	-2,012.6%	1,895.7%
Utility										
ENAGAS	ENAG		6,095	1%	12.1%	5.5%	24.3%	49.9%	-2,962.6%	3,700.4%
GAS NATURAL SDG	CTG		23,326	4%	7.1%	0.0%	27.9%	25.5%	-2,292.7%	3,052.3%
IBERDROLA	IBE		35,762	6%	9.8%	0.0%	29.7%	33.1%	-3,144.2%	4,699.8%
RED ELECTRICA CORPN.	REE		9,274	2%	17.9%	11.6%	24.5%	73.2%	-2,177.9%	3,666.2%
TELEFONICA	TEF		55,773	10%	2.9%	0.0%	23.4%	12.2%	-2,274.2%	2,994.7%
Transportation										
INTL.CONS.AIRL.GP.	IAG	24/01/11	9,611	2%	15.3%	0.0%	35.0%	43.8%	-1,943.0%	2,119.3%
Bank/Savings & Loan										
BANCO DE SABADELL	BSAB		9,407	2%	1.4%	0.0%	29.3%	4.9%	-1,911.1%	4,567.2%
BANCO POPULAR ESPANOL	POP		10,186	2%	-7.7%	0.0%	36.6%	-20.9%	-3,079.4%	5,169.3%
BANCO SANTANDER	SCH		91,241	16%	8.5%	0.0%	34.2%	24.9%	-2,985.3%	5,804.2%
BANKIA	BKIA	20/07/11	17,023	3%	-28.5%	-6.9%	141.1%	-20.2%	-12,857.1%	47,584.7%
BANKINTER 'R'	BKT		6,037	1%	11.2%	0.0%	36.7%	30.7%	-2,016.5%	3,626.2%
BBV.ARGENTARIA	BBVA		56,697	10%	5.1%	0.0%	34.0%	14.9%	-3,195.0%	5,507.0%
CAIXABANK	CABK	10/10/07	27,243	5%	6.9%	0.0%	32.8%	20.9%	-2,598.2%	4,233.1%
Insurance										
MAPFRE	MAP		8,635	2%	9.4%	0.0%	34.5%	27.3%	-3,145.1%	4,278.1%
Other Financial										
BOLSAS Y MERCADOS ESP	BOLS	14/07/06	2,524	0%	4.9%	0.0%	30.7%	15.9%	-1,952.9%	3,198.8%

* In Millons.

** Annualized return and volatility.

6 The optimization is implemented through Coordinate Descent algorithm

The statistical significance of the previous estimation is provided in the second step of the procedure by implementing Monte Carlo simulations as follows. Once $\Omega_{G-Lasso}$ is estimated, I move on to compute ρ in accordance to equation (1). Next, I check whether the elements of ρ are statistically different from zero for a pre-specified level of confidence α. To quantitatively determine the significance of each ρ_{ij}, a non-parametric kernel density of returns for each stock in the sample is adjusted.[7] Then, a sample of size 1.000 is independently drawn from each one of those densities. With this artificial and non-correlated sample of returns, a new partial covariance matrix is computed serving as a null hypothesis for the inference. I do this many times in order to obtain an empirical distribution for each ρ_{ij}. Finally, I calculate the percentiles for each of those distributions in accordance with a specified confidence level equal to α. When the original estimation of ρ_{ij} is larger than the percentile just mentioned, such ρ_{ij} is stated as statistically significant and retained as a link between node i and j, otherwise the connection between those nodes is discarded.

This 2-step estimation algorithm posits several benefits compared to more traditional applications. Among them it worth mentioning that it is a totally data-driven procedure in which neither topological constraints nor ad-hoc thresholds are imposed except for α. Appendix B provides evidence that supports its benefits by proving the result of its implementation with an artificially created dataset. In what follows, the *PCSNs* are constructed in accordance with this methodology.

7 A Gaussian kernel with smoothing parameter equal to 0.2 is assumed.

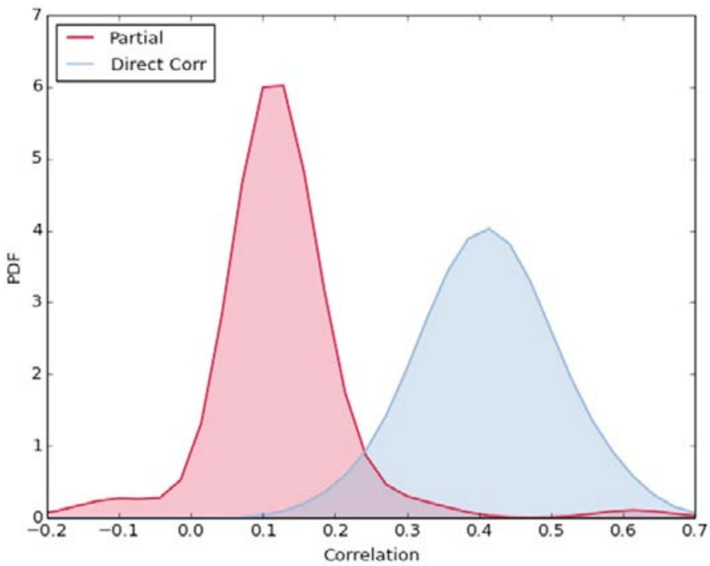
5 Empirical Results

The subsequent analysis is divided in two parts. The first one accounts for a static study by considering only the 250 most recent trading days (approximately one year). The aim is to describe the current state of the Spanish *PCSN*. The second part considers a dynamic perspective of the *PCSN* and therefore larger time series are required. In order to fulfill this data requirement, those firms whose inception date falls after the beginning of the dataset are excluded from the analysis. The parameters for the 2-step estimation algorithm are set as follows: 500 repetitions in order to obtain the an empirical distribution for each ρ_{ij}, a significance level of 1% and 10 fold Cross-Validation.

5.1 Static analysis

The results from the estimation reveal an optimal regularization parameter γ equal to 0.042. Figure 1 plots the non-parametric distribution and key descriptive statistics for the non-zero partial correlations for this optimal γ. In addition, the distribution of direct correlations estimated by means of optimal shrinkage Ledoit and Wolf (2004) is included for comparison.

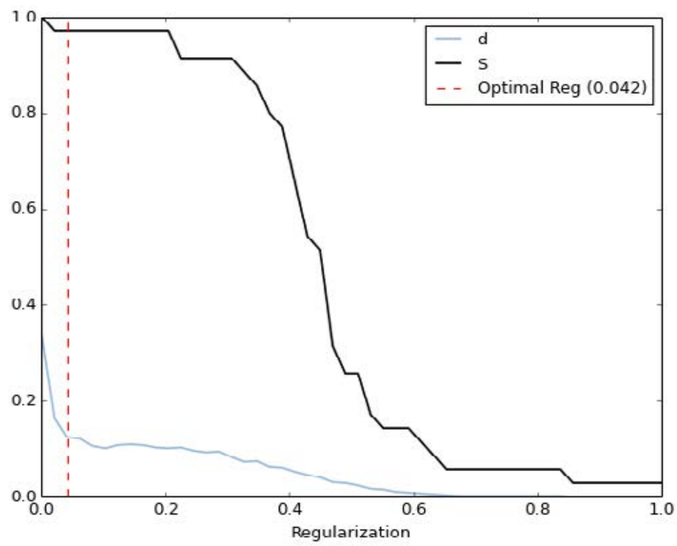
Non-parametric density of partial and direct correlation FIGURE 1



The direct correlation is a symmetric bell-shaped curve that resembles a normal density function. Relative to the partial correlation distribution, it shows fatter tails with a significantly lower mean value (0.412 vs 0.122). This is reasonable since the latter discards the co-movement of the returns processes due to the common market factor. Additionally, note that the direct distribution lies only on the positive quadrant while the partial correlation distribution shows negative values as well. The full partial correlation matrix and the direct correlation matrix are provided in appendix C. From the full matrix of partial correlation, it worth highlighting the extremely large value assumed by some of its components related to the banking and utility sectors (for the pair (BBVA, SCH) is 0.6. for (BSAB, POP) is 0.3 and (ENAG, CTG) is 0.3).

Figure 2 shows the density d and the size of the largest connected component d of networks corresponding to different levels of the regularization parameter γ.[8] There are three aspects that deserve some attention. First, S remains positive even for large values of γ evidencing an extremely persistent group of stocks intensively interrelated. This cluster of stocks belongs to the Banking sector as it is clearly seen in figure 3 below. Second, there is a non-linear and decreasing relationship between the γ and d. More interestingly, the optimal γ (vertical red line) roughly coincides with a turning point of d at around 0.15 evidencing a sharp transition between regimes with different slopes. Finally, note that *PCSN* never becomes connected unless γ is set to zero. Therefore, there exists a group of companies that are totally disconnected from the structure since they show non-significant partial correlation with the rest of the firms in the market.

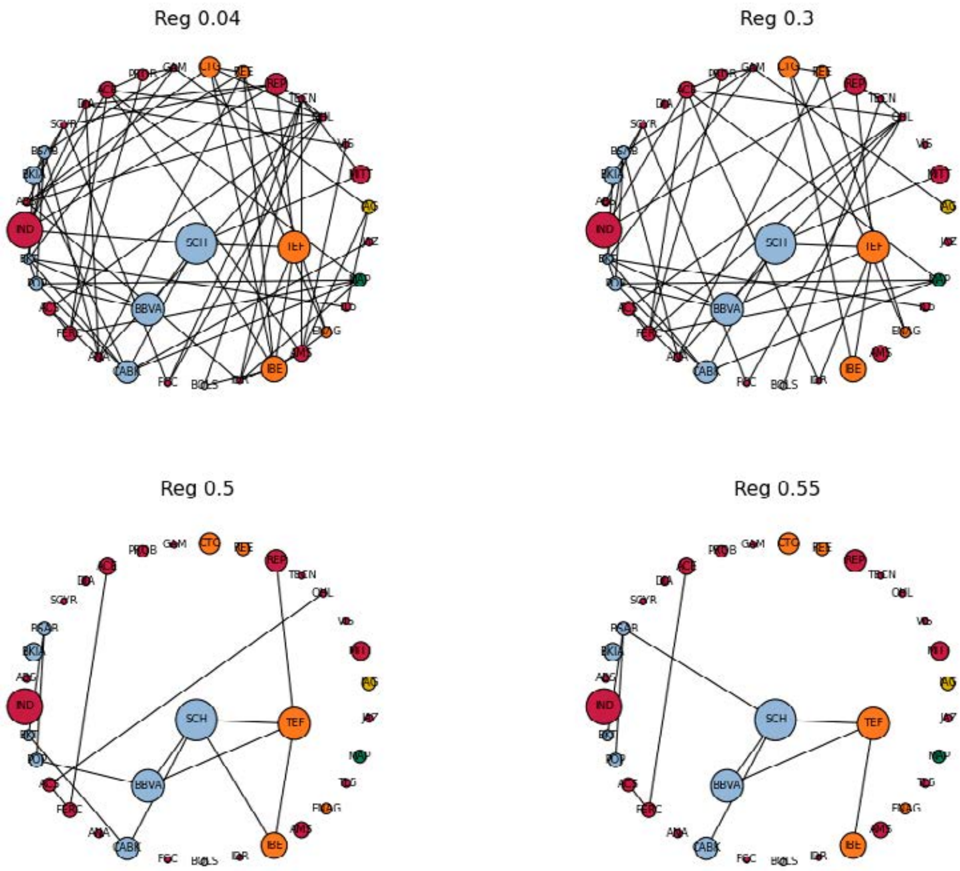
Network's measures for different levels of regularization FIGURE 2



The *PCSN's* for selected values of γ are depicted in figure 3 where the first one corresponds to the optimal *PCSN*, defined as the one associated with the optimal γ. The size of the nodes/stocks in the networks accounts for their (rescaled) market value for the last day in the sample. Their colors account for different industries (see figure caption for more details).

8 The larger the regularization parameter, the larger the weight on the L1-norm restriction which in turn increases in the sparseness in the resulting PCSN.

Networks for selected levels of regularization FIGURE 3



The colors of the nodes correspond to sectors: Industrial (Red), Bank/Savings & Loan (Blue), Utility (Orange), Insurance (Green), Other Financial (Grey) and Transportation (Yellow). The size of each node corresponds to its (rescaled) market value at Sep30, 2014.

The importance of the Banking sector is noticeable in figure 3. An artificial increase in γ eliminates less significant partial correlations. However, the cluster of nodes belonging to such sectors remains connected (e.g. SCH, BBVA, CABK, BSAB, POP). This gives us some insights about their *systemic importance*. Following the same reasoning, stocks from the utility sector also assume prominent positions in the structure (see the cases of TEL and IBE). On the other hand, the more peripheral role of the industrial stocks stems from their earlier disconnection from the network, as long as γ increases. This happens in spite of the large market capitalization shown by some of them (e.g. Inditex). A plausible explanation comes from the logic behind the functioning of the Banking sector. Its business is based on the interaction with other companies by providing financing and thus can take root in almost any segment of the economy. The next subsection is devoted to deeper analysis of the centrality of different economic sectors.

Key network measures are reported in table 2. The first four columns account for the networks from figure 3. The last column reports the average of the same set of measures over 1.000 realizations of networks made at random (Erdös-Rényi network) matched to the density of the optimal one. The evidence supports significant differences between the optimal Spanish *PCSN* and its random counterpart in several dimensions. Despite the fact that the largest connected component is barely smaller for the Spanish PCSN (97% vs 99%), such difference is statistically significant at conventional levels. As a consequence, its disconnected component is significantly larger than a random arrangement (3% vs 1%). This feature confers a sort of protection since the cluster of connected nodes is reduced preventing broader propagation of an initial shock. The large and significant value of transitivity for the Spanish *PCSN* relative to a random network (21% vs 13%) constitutes another potential source of strength. This is the case since conditioning on the network density, producing a large number of triangles, prevents local shocks becoming global by containing the former inside a group of tightly interrelated stock. Finally, it is worth mentioning that the Spanish *PCSN* seems to be disassortative (-0.17); highly connected stocks tend to be linked to poorly connected ones. This could be interpreted as a sign of weakness since any negative shock in poorly connected stock might be transmitted to rest of the market throughout its path toward the hub of the network. However, this effect is not statistically significant at conventional levels.

Network's measures for different levels of regularization* TABLE 2

	Optimal	0.30	0.5	0.55	Random Network
Basics					
Nodes	35	35	35	35	35
Links	77	48	15	8	76.9
Density	0.13	0.08	0.03	0.01	0.13
Mean Degree	4.40	2.74	0.86	0.46	4.40
Distance					
Diameter	5 (0.21)	7	4	3	4.989
Mean Distance	2.50 (0.43)	3.41	2.00	1.80	2.49
Components					
Largest Comp	0.97 (0.06)	0.91	0.26	0.14	0.99
Isolates	0.03 (0.05)	0.09	0.63	0.69	0.01
Patterns of Connectivity					
Transitivity	0.21 (0.00)	0.25	0.11	0.00	0.13
Assortativity	-0.17 (0.15)	-0.07	0.10	-0.56	-0.06

* Measures corresponding to the weigthed version of the PCSN.

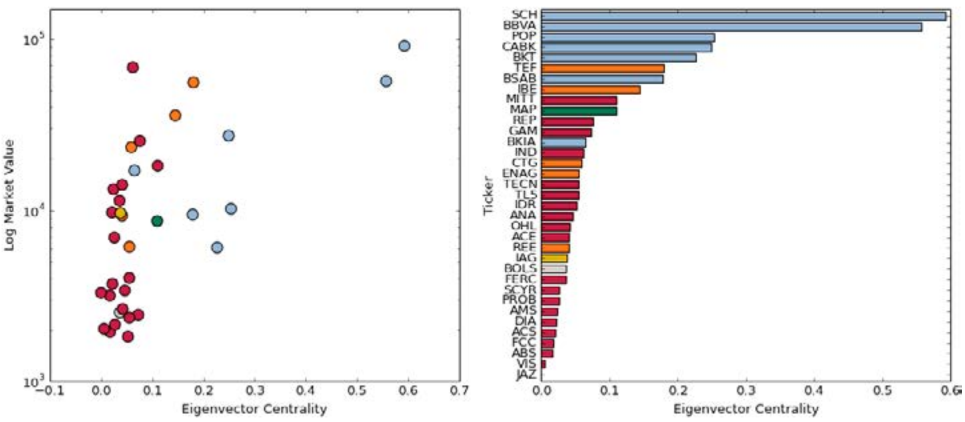
In parenthesis the p-values of the measures relative to the random counterpart.

5.1.1 Systemic Stocks in the Spanish Market

The *systemic importance* of a firm assesses its potential to greatly affect the performance of the entire market. In what follows I measure the systemic importance of stocks by means of the places they occupy in the Spanish *PCSN* which in turn are quantified by the eigenvector centralities, see section 3. Therefore, those stocks centrally located inside the *PCSN* rank high as systemic firms and as a consequence they could potentially cause large market movements.

The left-hand panel of figure 4 shows a scatter plot between a firm's centrality and its market capitalization (logarithmic scale) for the last day in the sample. The colors of the dots correspond to different industrial sectors (see the legend's figure for more details). The right-hand panel of the same figure depicts the centrality of each stock with their corresponding tickers (see table 1).

Relationship between centrality, market capitalization and economic sectors FIGURE 4



The colors of the nodes correspond to sectors: Industrial (Red), Bank/Savings & Loan (Blue), Utility (Orange), Insurance (Green), Other Financial (Grey) and Transportation (Yellow).

From figure 4, the positive relationship between market capitalization and stock centrality is evident. This pattern is also clear in table 3 which averages the same information at sector level. Additionally, note that the Banking sector is by far the most systemic one, followed by Insurance (with only one firm) and Utility sectors. It is particular interesting to note the case of the two most central firms in the market, Banco Santander and BBVA, showing a clear distinct pattern from the rest. Given this implicit hierarchy, a convenient macroprudential regulation should use such evidence with the aim of promoting a stable and sound stock market. Appendix D provides the full list of firms with their corresponding centrality score.

Mean and total centrality by economic sectors TABLE 3

		Eigenvector Centrality		Market Value*	
Industry	**Number of Firms**	**Total**	**Mean**	**Total**	**Mean**
Other Financial	1	0.037	0.037	2,524.4	2,524.4
Transportation	1	0.038	0.038	9,611.1	9,611.1
Insurance	1	0.110	0.110	8,635.1	8,635.1
Utility	5	0.481	0.096	130,229.9	26,046.0
Industrial	20	0.812	0.041	199,877.9	9,993.9
Bank/Savings & Loan	7	2.125	0.304	217,833.3	31,119.0
Total	**35**	**3.602**	**0.103**	**568,711.5**	

* In Millons.

In order to investigate deeply the relationship between market capitalization and centrality, the cross-sectional regression (7) is estimated by OLS with the aim of separating the market capitalization effect from effects of economic sectors. In this expression, $Centrality_i$ and MV_i corresponds to the eigenvector centrality and market value of firm *i*, respectively. D_j represents dummy variables for economic sectors where *Ind={Insurance (Ins), Utility (Ut), Industrial (Ind), Transportation (Tra), Other Financial (OF)}*. Note that significant and negative (positive) coefficients indicate lower (higher) centrality relative to the banking sector. Table 4 reports the results of the estimation

$$Centrality_i = \beta_0 + \beta_1 \ln(MV_i) + \sum_{j \in Ind} D_j + e_i \qquad (7)$$

A positive and statistically significant coefficient for β_1 is found evidencing that firms with large market capitalization tends to present high centrality scores which is consistent with the left-hand panel of figure 4. Further, the coefficient for the economic sector dummy variables are all negative and statistically significant at conventional levels. Then, the large centrality score of the banking sector is in part explained by its high market capitalization but it is also inherent to its business.

Regression of centrality on market capitalization and economic sectors TABLE 4

β_0	β_1	β_{Ins}	β_{Ut}	β_{Ind}	β_{Tra}	β_{OF}	R^2 Adj	F (pvalue)
-0.156	0.047	-0.156	-0.205	-0.203	-0.248	-0.180	0.670	0.000
(-1.086)	(3.305)***	(-1.881)*	(-4.569)***	(-5.286)***	(-2.997)***	(-2.062)**		

Significance * p<0.10, ** p<0.05, *** p<0.01.

5.1.2 International evidence

In order to compare the Spanish *PCSN* with other structures, the same methodology is applied to the set of stocks constituting the market indexes FTSE-100, CAC-40, DAX-30 and FTSE-MIB as representatives of the UK, French, German and Italian market, respectively. I rely on Datastream to construct similar datasets as in the Spanish case covering the exact same time period. In table 5 a comparison between the network topologies of theses stock markets is provided.

Notice that each of the network presents a sizable largest connected component and also a high value of transitivity. The largest connected component of the UK network is somehow smaller 82% which could be interpreted as a source of strength as it was commented before. In terms of connectivity, the Spanish and German structured show the largest density (13%) while the UK network is the less connected one (3%). This feature is also evident in the mean distance where it is 2.50 for the case of Spanish *PCSN* while it is 5.00 for the British network. Note, moreover, that the only market showing a disassortative arrangement is Spain which could comprise a source of weakness, as was commented before.

Network's measures - International comparison TABLE 5

	Spain	UK	France	Germany	Italy
Basics					
Nodes	35	101	40	30	37
Links	77	136	70	58	66
Density	0.13	0.03	0.09	0.13	0.10
Mean Degree	4.40	2.69	3.50	3.87	3.57
Distance					
Diameter	5	11	6	7	7
Mean Distance	2.50	5.00	2.84	2.86	3.02
Components					
Largest Comp	0.97	0.82	0.93	1.00	0.97
Isolates	0.03	0.03	0.03	0.00	0.03
Patterns of Connectivity					
Transitivity	0.21	0.19	0.09	0.21	0.15
Assortativity	-0.17	0.04	-0.07	0.07	0.12

* Measures corresponding to the weigthed version of the PCSN.

The association between market capitalization and centrality is quantified in table 6. This table reports the estimation of equation 7 for each of the markets under analysis. The coefficient β_1 is positive and significant for all *PCSN* except the French case which is non-statistically significant. In general terms, the sectorial dummies tend to be negative and statistically significant for French, German and Italian *PCSN* as they are for the Spanish case evidencing the prominent role of the banking system. However, a distinctive feature in the French *PCSN* is that its Insurance sector is on average more central than the Banking sector in accordance to the positive and significant coefficient β_{Ins}. This result should be taken with caution since there is only one stock representing that sector. Interestingly, the UK market behaves in a totally different way since none of the sectorial dummy variables are statistically different from zero. Therefore, for the UK market, members of the banking sector do not necessarily take central positions in the network. This result could be driven by the low connectivity level found in the UK *PCSN* (3%).

In order to facilitate the comparison, figure 5 plots the mean eigenvector centrality at sector level for each of the countries in the sample. I normalize the eigenvector centralities by requiring that the sum of its components equals the number of stock in each market, thus accounting for different network sizes. The banking sector assumes a critical role in Spain, France and Italy.[9] For the German case, companies from the Utility sector show on average the highest central position. The UK network presents a distinctive behavior since the centrality in this market is roughly evenly distributed inside the industrial sector. These structural differences in the *PCSN* could shed some light on the differential response of European stock markets when facing the same shock. Further analysis is required in this regards. Appendix F provides descriptive tables of centrality and market capitalization for each country in the study.

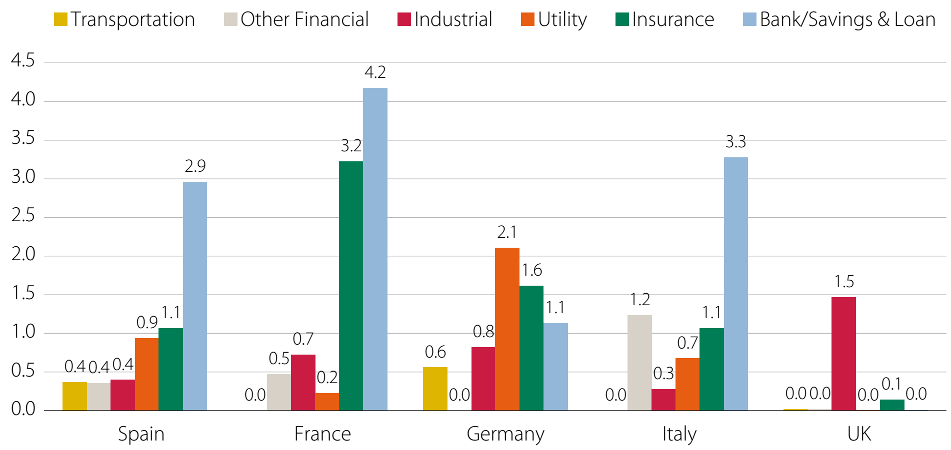
9 In this calculation, the increased centrality of the Insurance sector in France does not capture the evidence presented in table 6. Probably, this is due to the fact that this sector is represented by only one firm.

Regression of centrality on market capitalization and economic sectors - International Comparison TABLE 6

	Spain	UK	France	Germany	Italy
β0	-0.156 (-1.086)	-0.156 (-1.755)*	0.153 (1.307)	-0.247 (-1.538)	-0.038 (-0.417)
β1	0.047 (3.305)***	0.020 (1.836)*	0.000 (0.005)	0.048 (2.516)**	0.021 (1.682)*
βIns	-0.156 (-1.881)*	-0.025 (-0.814)	0.181 (3.011)***	-0.022 (-0.321)	-0.050 (-1.003)
βUt	-0.205 (-4.569)***	-0.040 (-1.045)	-0.133 (-3.099)***	0.066 (1.141)	-0.080 (-2.224)**
βInd	-0.203 (-5.286)***	0.006 (0.286)	-0.079 (-3.162)***	-0.110 (-2.837)***	-0.107 (-4.469)***
βTra	-0.248 (-2.997)***	-0.019 (-0.385)		-0.124 (-2.151)**	
βOF	-0.180 (-2.062)**	-0.018 (-0.551)	-0.101 (-1.825)*		0.028 (0.628)
R^2 Adj	**0.670**	**0.007**	**0.743**	**0.384**	**0.702**
F (pvalue)	**0.000**	**0.354**	**0.000**	**0.004**	**0.000**

Significance * p<0.10, ** p<0.05, *** p<0.01.

Centrality by economic sectors – International comparison FIGURE 5

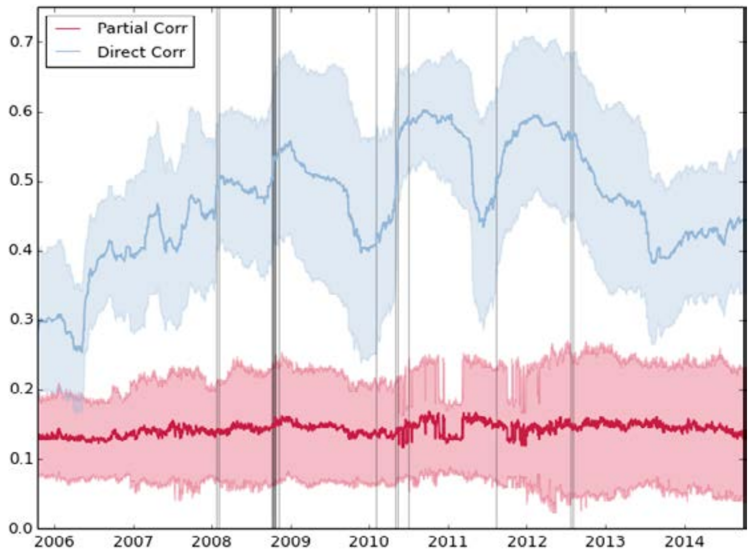


5.2 Dynamic analysis

This section studies the dynamical properties of the Spanish *PCSN* and discusses its stability through time by relying on a moving window approach as follows. The 2-step estimation algorithm (see section 4) is implemented upon moving windows of returns of 250 trading days of length. Since a one-day displacement step is considered, 2.336 *PCSN* are estimated with their corresponding network measures and firm's centrality scores. Since large time series of data are required in the section, those firms with inception dates later than Nov 1, 2004 (10 firms) are discarded for this analysis. I define *(negative) market events* as market index returns lower than three standard deviations computed for the entire sample period. The dates for such market events are plotted in the following figures as vertical black lines.

Let us start by describing figure 6 which plots the time series of the mean of non-zero partial correlations and the mean of direct correlations estimated using optimal shrinkage Ledoit and Wolf (2004). The colored regions account for the corresponding +/- one standard deviation away from the corresponding mean. As expected, the mean direct correlation shows three humps roughly coinciding with periods of high financial distress. The mean partial correlation presents a totally different behavior since its time evolution remains quite stable with average level around 0.12.

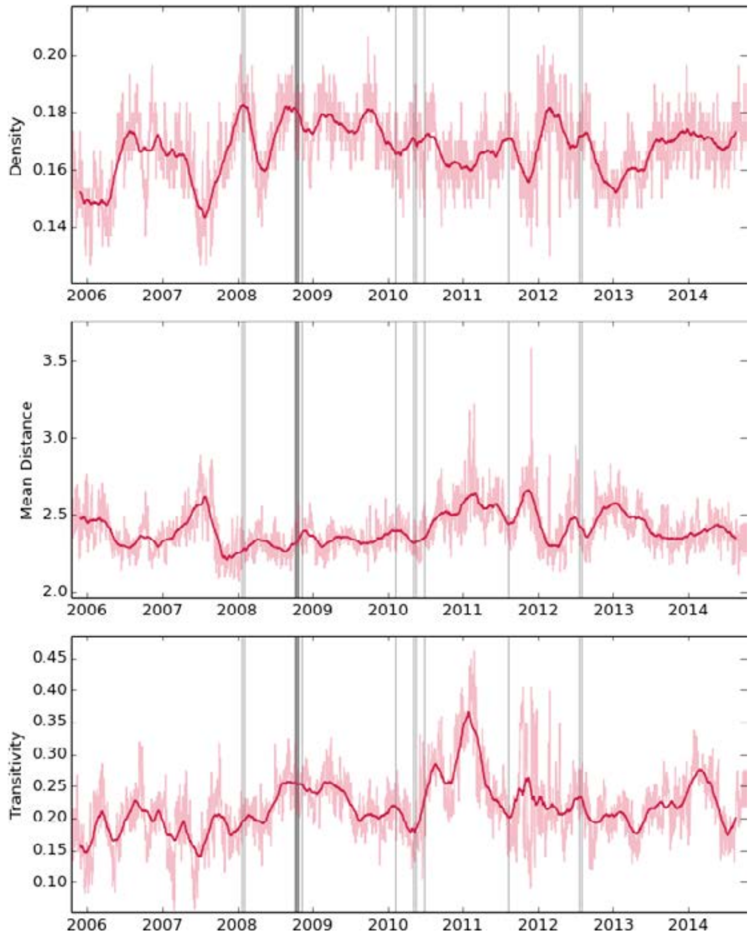
Time series of mean correlation and +/- one standard deviation FIGURE 6



The time series of three selected network measures, Density, Mean Distance and Transitivity, are depicted in figure 7. Light-red lines accounts for the exact quantities while the dark-red line corresponds to the 60-days centered moving average. Visual inspection does not allow us to conclude about any clear pattern in the data regarding the performance of those network measures around market events. Nevertheless, section 6 provides rigorous statistical assessment in this respect.

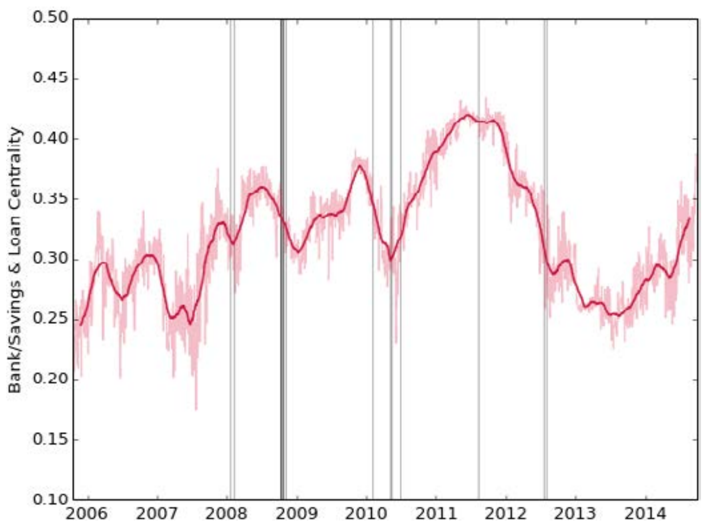
Despites the comments already mentioned about figure 7, a quite clear pattern arises when the centrality of the banking industry (mean of the centrality of the firms belonging to this sector) is under analysis. Figure 8 plots the time series of the mean centrality of the banking sector in light-red and its 60-days center moving average in dark-red. Note that negative market shocks tend to occur when the centrality of the banking sectors presents a downward slope. In other words, before a market event, the centrality of the banking sectors reaches a peak and starts to decline afterwards. In fact, there is just one episode in mid-2010 for which this observation is not true. As it was mentioned, section 6 investigates this phenomenon further.

Time series of selected network's measures FIGURE 7



Network Measure through time in light-red and the related 60 day centered moving average in dark-red.

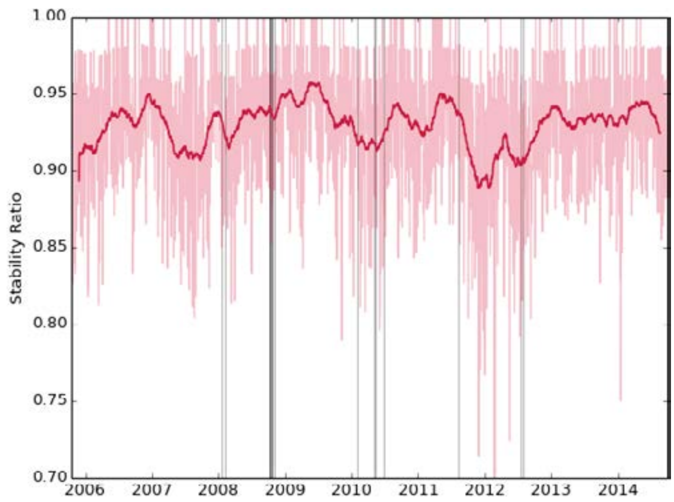
Banking sector centrality through time FIGURE 8



A final comment regards to the stability of the *PCSN* through time. Since each of the *PCSN* is composed of exactly the same set of stocks/nodes, the only elements that change from period *t-1* to *t* are the links in subsequent networks. Therefore, I will measure the stability of the Spanish *PCSN* as the proportion of active links in period *t-1* that remains active in *t*. To be more concrete, let us define the set of *PCSN* indexed by time as $\psi = \left\{\varphi_1^u, \varphi_2^u, \ldots, \varphi_{2337}^u\right\}$ and their corresponding set of links as $\Phi = \left\{\varrho_1, \varrho_2, \ldots, \varrho_{2337}\right\}$. The *intersection network* in period *t*, $\tilde{\varphi}_t^u$, is built upon the intersection of two subsequent set of links as $\tilde{\varrho}_t = \varrho_{t-1} \cap \varrho_t$. Then the stability ratio SR_t in period *t* is given by formula (8) and its time series is plotted in figure 8.

$$SR_t = \frac{\#\, of\ elements\, in\ \tilde{\varrho}_t}{\#\, of\ elements\, in\ \varrho_{t-1}} \tag{8}$$

Stability ratio through time FIGURE 9

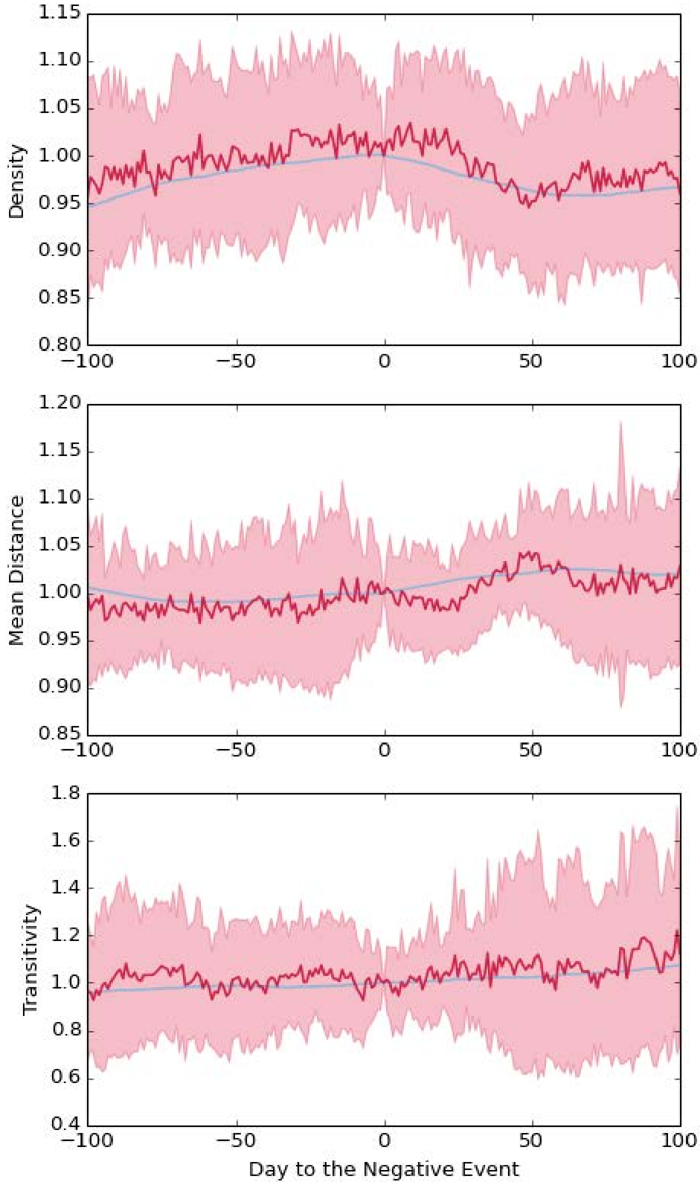


As before, the light-red lines correspond to the exact stability ratio in period *t* while the dark-red line accounts for its 60-days centered moving average. The mean and standard deviation of the centered SR_t is 93% and 4%, respectively. The fact that more than 90% of the links remains active from subsequent networks confers a sort of structural stability through time to the Spanish *PCSN*. Note also that the minimum values of SR_t are reached in the periods following the inclusion of Bankia as a new constituent of the IBEX-35, say after Ago-2011. In fact, the centered SR_t shows a negative trend starting in Ago-2011 that reverts by the end of that year (values of this variable below 90% corresponds to the period Oct-2011 until Feb-2012). In this regard it could be said that the market required approximately 6 months to digest such network reconfiguration.

6 Network measures as leading indicators of market instability

The extent to which a set of network measures might be used as leading indicators of market distress is investigated in this section. The behavior of three network-based measures, Density, Mean Distance and Transitivity, around negative market events is depicted in figure 10.

Behavior of network's measures around a market event FIGURE 10



In the above figure market events are aligned at day zero and 100 trading days before and after such an event are considered. This alignment is plotted in the x-axis. The y-axis shows the mean of each network-based measure in dark-red, the mean of its corresponding 60-days length moving average in blue and the region comprising +/- one standard deviation away from mean in light-red. In order to account for different levels of the network measure when a market event takes place, I normalize such measures to 1 in such dates.

From figure 10 there is no identifiable pattern for the transitivity. However, this is not the case for the density and the mean distance. Considering the network density, let us note that it tends to be above its mean value at the moment of the negative shock for at least 30 trading days before such event. For the mean distance, the reverse is true since about 30 trading days before a large and negative market movement this measure tends to be below its mean value at the moment of the negative shock. Therefore, it seems that the Spanish *PCSN* presents particular features that seem to anticipate market distress episodes. In particular it could be said that just before a negative event, the stock network becomes densely connected while shortening its mean distance. This symmetric behavior is expected since these two variables shows a strong negative correlation, see appendix E.

In order to better understand this phenomenon two alternative and complementary approaches are pursued. Subsection 6.1 shows the results of the estimation of a Probit model where dependent variable assumes values equal to one when the market suffers a negative return larger than three standard deviations (market event) and zero otherwise. In this specification, the set of independent variables are the lagged network measures from figure 10. In subsection 6.2, it is assumed that the daily return process follows an *ARCH* model where the equation of the conditional variance includes as independent variables the same (lagged) network measures as in the previous specification.

6.1 A Probit model for negative market movement with network measures as independent variables

I estimate a Probit model to quantitatively assess the extent to which lagged network-based measures can predict large and negative market movements. The dependent variable is dichotomous assuming value one if the market suffers from a negative return larger than three standard deviations (computed from the entire dataset) and zero otherwise. As explanatory variables I include the two of the network measures depicted in figure 10, Density and Transitivity[10]. Additionally, the centrality of the banking sector is also introduced as an independent variable. Since predictability is at the center of the analysis, I consider lags at 10 and 30 trading days in setting the explanatory variables. Given that the set of regressors shows a significant level of autocorrelation (see appendix E), standard errors tend to be downward biased leading to incorrect t-statistics. In order to account for the correction in the standard errors due to heteroskedasticity and autocorrelation, I implement bootstrap methods as suggested by Berg and Coke (2004).

10 I discard the mean distance as an additional independent variable given the high negative correlation between this variable and density (see Appendix E).

Table 7 reports the results of the estimation of three different models that progressively include the independent variables. Model 1 presents positive and significant effects of network density on the probability of a large and negative market movement. Therefore, a dense PCSN shows a higher probability to suffer a large negative return, an observation consistent with the evidence presented in figure 10. The inclusion of the banking centrality as an additional independent variable in model 2 also shows a positive and significant effect at lag 30 while retaining the significance of network density. Therefore, a network more densely connected in which the banking sector assumes a central position is consistent with an increased probability for a large and negative shock in the market. Finally, in model 3, all of the considered network measures are included in the estimation. However, transitivity does not show any significant coefficient for any of its lags, an observation which is also consistent with figure 10. In any case, the conclusions regarding the signs of the coefficients and their significance for the case of density and banks centrality remain valid.

Results of Probit model's estimations TABLE 7

Density	**Model 1**	**Model 2**	**Model 3**
Lag 10	21.23 (2.2)**	18.32 (1.77)*	18.81 (1.83)*
Lag 30	21.94 (2.37)**	25.14 (2.27)**	24.93 (2.28)**
Banks Centrality			
Lag 10		-6.52 (-1.22)	-6.17 (-1.12)
Lag 30		10.27 (2.02)**	10.00 (1.85)*
Transitivity			
Lag 10			-1.22 (-0.53)
Lag 30			0.57 (0.22)
Likelihood Ratio Index	**0.099**	**0.127**	**0.129**
LLR p-value	**0.000**	**0.000**	**0.000**

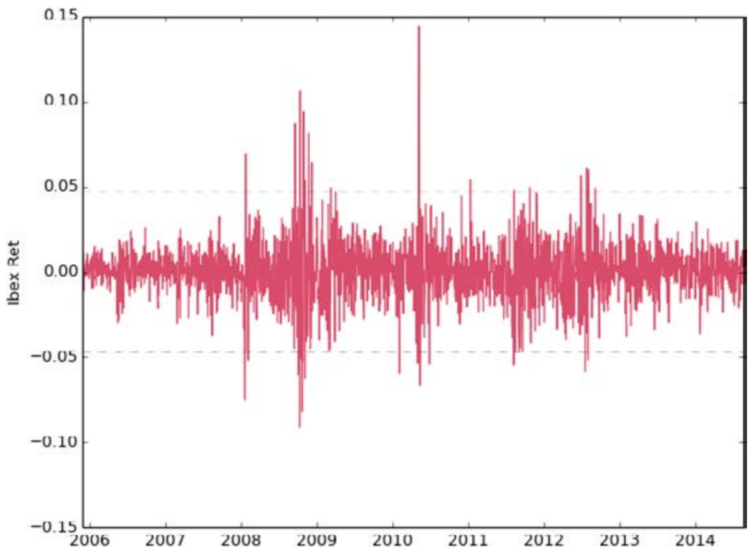
SE estimated by Bootstrapping. Significance * p<0.10, ** p<0.05, *** p<0.01.

In summary it could be said that there is some evidence that network-based measures are a leading indicator of large market distress events. In particular, the statistical results show that the probability for a large negative movement increases when Spanish *PCSN* becomes more densely connected. In cases where the banking sector assumes more central position in the structure, the effect is reinforced by increasing further the probability of this type of episodes.

6.2 ARCH models for the Spanish Market with network measures as independent variables

The heteroskedasticity of the return process is a well-documented phenomenon in the econometric literature, Hamilton (1994). Figure 11 plots the time series IBEX-35 return where such empirical regularity is observed. The periods mid 2008-2009 and 2012-2013 are characterized by a high volatile regime while the period 2006-2008 shows a tranquil market.

Time series of daily market returns (IBEX-35) FIGURE 11



Following the tradition in the econometric literature, (see Bollerslev (1987) and Hamilton (1994)) I assume that return process is described by an *ARCH(q)* model specified in equations (9) to (11) where the error term η_t comes from a t-student distribution with ν degrees of freedom. The novelty of this specification arises from the inclusion in the equation of the conditional variance σ_t^2 (equation 11) lagged network measures as explanatory variables. As in the previous subsection, the set of lagged network measures considered for the experiment is *N={Density, Transitivity, Centrality of the Banking Sector}*.

$$r_t = c_r + \varepsilon_t \tag{9}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{10}$$

$$\sigma_t^2 = c_\sigma + \sum_i^q \alpha_i \varepsilon_{t-i}^2 + \sum_{n \in N} \sum_{j \in \{10,30\}} \alpha_j^n \, n_{t-j} \tag{11}$$

The model (9) to (11) is estimated by maximizing the log likelihood where ν is considered as an additional parameter. The order of the process is set to be equal to 5 since in non-tabulated results, the coefficients α_i beyond that period are not statistically significant. Table 8 shows the estimated parameter for three models with the sequential inclusion of Density, Centrality of the Banking Sector and Transitivity as independent variables.

Model 1 considers the lags of density as the only network-based independent variables. In this model, lag 10 shows a positive and statistically significant coefficient which means that denser networks increased the conditional market volatility. In model 2, the centrality of the banking sector is considered as an additional independent variable. A positive and statistically significant coefficient is found for its lag 30. Moreover, the sign and significance of the density at lag 10 remains. The comparison between model 1 and 2 from table 7 with those models from table 8 confers some robustness on the results. Therefore, denser networks where the centrality of the banking sector is high show larger probabilities of a strong negative movement which could be explained by their positive effect of those variables on the conditional market volatility.

Model 3 in table 8 presents the major difference with respect to table 7. In the former table, the transitivity level at lag 30 shows a negative and significant effect on market variance. At first this evidence might seem to be counterintuitive, but as discussed in Newman and Park (2003), large transitivity is consistent with networks arranged in communities. In other words, an increase in the number of triangles in the network, controlling by its density, leads to an internal organization of the structure characterized by groups of stocks tightly interrelated. This would undermine the possibility of macro effect as a consequence of shocks at micro level and thus reducing the conditional market variance.

Results of ARCH model's estimations TABLE 8

Density	Model 1	Model 2	Model 3
Lag 10	14.67	9.64	12.83
	(4.12)***	(2.76)***	(3.21)***
Lag 30	2.91	-1.21	-0.85
	(0.71)	(-0.29)	(-0.22)
Banks Centrality			
Lag 10		0.40	1.24
		(0.33)	(0.89)
Lag 30		3.73	3.82
		(3.12)***	(2.92)***
Transitivity			
Lag 10			0.23
			(0.23)
Lag 30			-2.33
			(-2.51)**
Squared Residuals			
Lag 1	0.08	0.07	0.07
	(2.77)***	(2.51)**	(2.45)**
Lag 2	0.14	0.13	0.12
	(4.17)***	(3.94)***	(3.91)***
Lag 3	0.16	0.15	0.14
	(4.82)***	(4.63)***	(4.55)***
Lag 4	0.21	0.19	0.18
	(5.77)***	(5.6)***	(5.38)***
Lag 5	0.16	0.16	0.15
	(4.51)***	(4.51)***	(4.51)***
Others Parameter			
T-Student df	6.33	6.64	6.69
	(6.51)***	(6.39)***	(6.28)***
Akaike criterion	**3.427**	**3.422**	**3.421**
Schwarz criterion	**3.452**	**3.452**	**3.456**
Log Likelihood	**-3,893.7**	**-3,885.5**	**-3,882.7**

Significance * p<0.10, ** p<0.05, *** p<0.01.

7 Conclusion and Future Research Lines

Network theory as a general analytic framework has been considered extremely useful in several branches of sciences. Recently, the financial research community has started to adopt it in order to get new and valuable insights in the context of stock markets. The current paper proposes a 2-step algorithm to estimate the Spanish Partial Correlated Stock Network. This algorithm is designed to overcome or at least to attenuate the shortcomings of standard techniques. The estimated network is formed by the constituents of IBEX-35 as nodes and statistically significant partial correlations as links connecting pairs of them. Once the network is in place, different network-based measures are computed.

Three major results come from the empirical study. First, consistent with general wisdom, the banking sector assumes a central position in the network and thus could severely affect the entire market. Particularly interesting is the case of Banco Santander and BBVA given their extremely high centrality scores. Therefore, an appropriate macroprudential regulation scheme should accommodate this evidence in order to consider the differential effect of fundamental market players. Second, the investigated network is found to be quite stable over time and presents properties that are inconsistent with a random arrangement. It is worth mentioning that the largest connected component is smaller, and the transitivity level is larger in comparison with those features coming from Erdös-Rényi networks. International comparability is also addressed by identifying common traits between the *PCSN* from Spain, UK, France, Germany and Italy. All of such networks present sizable values for the largest connected component and transitivity. However, the Spanish and German structures are the more densely connected while the UK network is extremely sparse. The central positions occupied by different economic sectors in those structures reveal a distinctive organization of such networks which prevents simple comparisons. Finally, there is also evidence that the current state of the Spanish *PCSN*, captured by key network measures, could be used as leading indicators of market distress. Therefore, including a set of network-based measures as part of the early-warning tool box for market regulators seems to be a suitable choice in order to properly assess systemic risk.

Two main research lines are left for future studies. The first one consists of an in-depth analysis of the similarities and differences between the largest stock markets around the world based on this unified framework. Such methodology could shed some light on the heterogeneous performances of different countries across the recent financial crisis. Another promising research line relates to the consideration of a directed stock network instead of the undirected one which is analyzed in the current paper. The possibility of including directed links in the structure would allow us to enrich the framework and to complement the conclusions already provided in this paper. For example, it might be possible to differentiate those stocks that tend to initiate a shock, *"threatening stocks",* from those that tend to receive it, *"vulnerable stocks"*. Therefore, the study of stock markets as directed networks should not be considered as a mere intellectual curiosity but instead as a valuable framework with the potential to enhance our understanding on the nature of the firms as parts of the entire market.

References

Allen, F., and A. Babus, 2009, Networks in Finance, *In: Kleindorfer, P., Wind, J. (Eds.), The Network Challenge: Strategy, Profit, and Risk in an Interlinked World, Wharton School Publishing*, 367–382.

Barabási, A. L., and R. Albert, 1999, Emergency of Scaling in Random Networks, *Science* 286, 509–512.

Berg, A., and R. N. Coke, 2004, Autocorrelation-Corrected Standard Errors in Panel Probits: An Application to Currency Crisis Prediction, *IMF Working Papers* 4, 1–18.

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

Bollerslev, T., 1987, A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *The Review of Economics and Statistics* 3, 542–547.

Bonacich, P., 1972, Factoring and weighting approaches to status scores and clique identification, *The Journal of Mathematical Sociology* 2, 113–120.

Bonacich, P., 2007, Some unique properties of eigenvector centrality, *Social Networks* 29, 555–564.

Bonanno, G., G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, and R. N. Mantegna, 2004, Networks of equities in financial markets, *The European Physical Journal B - Condensed Matter* 38, 363–371.

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

Diebold, F. X., and K. Yılmaz, 2014, On the network topology of variance decompositions: Measuring the connectedness of financial firms, *Journal of Econometrics* 182, 119–134.

Freeman, L., 1978, Centrality in Social Networks Conceptual Clarification, *Social Networks* 1, 215–239.

Friedman, J., T. Hastie, and R. Tibshirani, 2008, Sparse inverse covariance estimation with the graphical lasso., *Biostatistics (Oxford, England)* 9, 432–41.

Garas, A., and P. Argyrakis, 2007, Correlation study of the Athens Stock Exchange, *Physica A: Statistical Mechanics and its Applications* 380, 399–410.

Haldane, A., 2009, Rethinking the financial network, (Speech delivered at the Financial Student Association, Amsterdam, April.).

Hamilton, J. D., 1994, *Time Series Analysis* (Princeton University Press).

Hamilton, J. D., 1994, Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* 64, 307–333.

Hautsch, N., J. Schaumburg, and M. Schienle, 2014, Financial Network Systemic Risk Contributions, *Review of Finance*.

Hautsch, N., J. Schaumburg, and M. Schienle, 2014, Forecasting systemic impact in financial networks, *International Journal of Forecasting* 30, 781–794.

Huang, W. Q., X. T. Zhuang, and S. Yao, 2009, A network analysis of the Chinese stock market, *Physica A: Statistical Mechanics and its Applications* 388, 2956–2964.

Jackson, M. O., 2010, *Social and Economic Networks* (Princeton University Press).

Jackson, M. O., 2014, Networks in the Understanding of Economic Behaviors, *Journal of Economic Perspectives*, 1–23.

Jung, W., S. Chae, J. S. Yang, and H. Moon, 2006, Characteristics of the Korean stock market correlations, *Physica A: Statistical Mechanics and its Applications* 361, 263–271.

Ledoit, O., and M. Wolf, 2004, Honey, I Shrunk the Sample Covariance Matrix, *The Journal of Portfolio Management* 30, 110–119.

Mantegna, R. N., 1999, Hierarchical structure in financial markets, *The European Physical Journal B - Condensed Matter* 11, 193–197.

Milgram, S., 1967, The Small-World Problem, *Psychology Today*, 60–67.

Newman, M. E. J., 2003, The structure and function of complex networks, *SIAM review* 45, 167–256.

Newman, M. E. J., 2004, Analysis of weighted networks, *Physical Review E*.

Newman, M. E. J., 2010, *Networks: An Introduction Oxford University Press* (Oxford University Press).

Newman, M. E. J., and J. Park, 2003, Why social networks are different from other types of networks, *Physical Review E* 68.

Onnela, J. P., A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, 2003, Asset trees and asset graphs in financial markets, *Physica Scripta*, 48–54.

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

Ozsoylev, H., J. Walden, and M. D. Yavuz, 2014, Investor Networks in the Stock Market, *Review of Financial Studies* 27, 1323–1366.

Peralta, G., and A. Zareei, 2014, A Network Approach to Portfolio Selection *SSRN - Working Paper*.

Pozzi, F., T. Di Matteo, and T. Aste, 2013, Spread of risk across financial markets: better to invest in the peripheries, *Scientific reports* 3, 1665.

Schweitzer, F., G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White, 2009, Economic networks: the new challenges., *Science* 325, 422–5.

Stevens, G., 1998, On the Inverse of the Covariance in Portfolio Analysis, *The Journal of Finance* 53, 1821–1827.

Tse, Chi K., J. Liu, and F. C. M. Lau, 2010, A network perspective of the stock market, *Journal of Empirical Finance* 17, 659–667.

Tumminello, M., F. Lillo, and R. N. Mantegna, 2010, Correlation, hierarchies, and networks in financial markets, *Journal of Economic Behavior & Organization* 75, 40–58.

Ugander, J., B. Karrer, L. Backstrom, and C. Marlow, 2011, The Anatomy of the Facebook Social Graph.

Vandewalle, N., F. Brisbois, and X. Tordoir, 2001, Non-random topology of stock markets, *Quantitative Finance* 1, 372–374.

Whittaker, J., 1990, *Graphical Models in Applied Multivariate Statistics* (John Wiley & Sons, Inc.).

Yellen, J. L., 2013, Interconnectedness and systemic risk – lessons from the financial crisis and policy implications, (Board of Governors of the Federal Reserve System, at the American Economic Association/American Finance Association Joint Luncheon Washington, DC.).

Appendix A

In this appendix I follow Stevens (1998) to show the connection between the partial correlations of returns and the regression coefficients resulting from regressing the return of a particular stock with the rest of them.

Let us assume that the correlation matrix of returns is given by the $n \times n$ matrix $\Sigma = \left[\sigma_{ij}\right]$ which could be partitioned as follows:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{1j} \\ \sigma_{j1} & \acute{O}_{n-1} \end{bmatrix} \tag{A.1}$$

where σ_{11} is the variance of returns for asset 1, σ_{1j} is a $1 \times n-1$ vector of covariance between stock 1 and the other $n-1$ stocks in the sample and Σ_{n-1} corresponds to the sub-matrix of Σ resulting from the elimination of its first row and column. Defining the inverse of the covariance matrix as $\Omega = \left[\omega_{ij}\right] \equiv \Sigma^{-1}$, its partitioned form is written as

$$\Omega = \begin{bmatrix} \omega_{11} & \omega_{1j} \\ \omega_{j1} & \Omega_{n-1} \end{bmatrix} \tag{A.2}$$

Since $\Omega^* \Sigma = I$, the next two expressions could be derived (first row of Ω with the columns of Σ)

$$\omega_{1j} = -\omega_{11}\sigma_{1j}\Sigma_{n-1}^{-1} = -\omega_{11}\sigma_{1j}\Omega_{n-1} \tag{A.3}$$

$$\omega_{11} = \left[\sigma_{11} - \sigma_{1j}\Sigma_{n-1}^{-1}\sigma_{j1}\right]^{-1} = \left[\sigma_{11} - \sigma_{1j}\Omega_{n-1}\sigma_{j1}\right]^{-1} \tag{A.4}$$

Note that the term $\sigma_{1j}\Sigma_{n-1}^{-1}$ in A.3 corresponds to the row vector of the regression coefficients that result from regressing stock 1 returns on the rest of the stocks. We call this vector $\beta_1^{'}$. Additionally, by definition R^2 in such regression is given by

$$R_1^2 \equiv \frac{\sigma_{1j}\Sigma_{n-1}^{-1}\sigma_{j1}}{\sigma_{11}} \tag{A.5}$$

Therefore, equation (A.4) is restated as

$$\omega_{11} = \left[\sigma_{11}\left(1 - R_1^2\right)\right]^{-1} = \frac{1}{\sigma_{e1}^2} \tag{A.6}$$

where σ_{e1}^2 accounts for the proportion of the total return's variance of stock 1 that is not explained by the regression or the variance of the residual. Finally, introducing expression A.6 into A.3

$$\omega_{1j} = -\frac{\beta_1^{'}}{\sigma_{e1}^2} \tag{A.7}$$

Although A.6 and A.7 assumes the dependent variable in the regression is the return of stock 1, a convenient permutation of row and columns allows us to fully characterize Ω as follows:

$$\Omega = \begin{bmatrix} \frac{1}{\sigma_{e1}^2} & -\frac{\beta_{12}}{\sigma_{e1}^2} & \cdots & -\frac{\beta_{1n}}{\sigma_{e1}^2} \\ -\frac{\beta_{21}}{\sigma_{e2}^2} & \frac{1}{\sigma_{e2}^2} & & -\frac{\beta_{2n}}{\sigma_{e2}^2} \\ \vdots & & \ddots & \vdots \\ -\frac{\beta_{n1}}{\sigma_{en}^2} & -\frac{\beta_{n2}}{\sigma_{en}^2} & \cdots & \frac{1}{\sigma_{en}^2} \end{bmatrix} \quad \text{(A.8)}$$

On the other hand, we have defined the matrix of partial correlation coefficients between stocks as

$$\rho = \left[\rho_{ij}\right] = -\Delta\Omega\Delta \quad \text{(A.9)}$$

where $\Delta = diag\left(1/\sqrt{\omega_{ii}}\right)$. Simple calculations show that

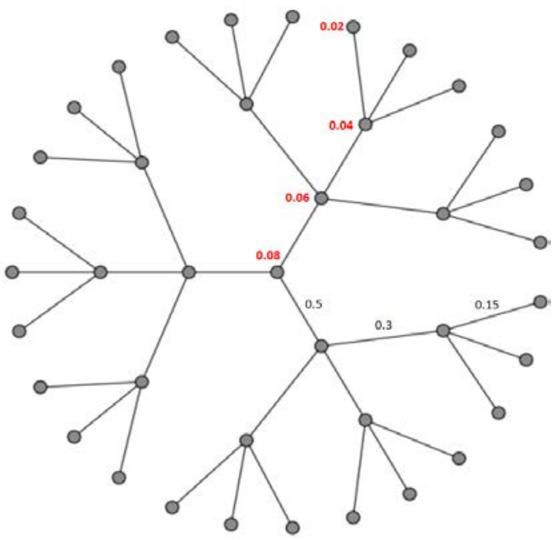
$$\rho = \begin{bmatrix} -1 & \beta_{12}\frac{\sigma_{e2}}{\sigma_{e1}} & \cdots & \beta_{1n}\frac{\sigma_{en}}{\sigma_{e1}} \\ \beta_{21}\frac{\sigma_{e1}}{\sigma_{e2}} & -1 & & \beta_{2n}\frac{\sigma_{en}}{\sigma_{e2}} \\ \vdots & & \ddots & \vdots \\ \beta_{n1}\frac{\sigma_{e1}}{\sigma_{en}} & \beta_{n2}\frac{\sigma_{e2}}{\sigma_{en}} & \cdots & -1 \end{bmatrix}. \quad \text{(A.10)}$$

Appendix B

This appendix is devoted to testing the estimation methodology explained in section 4 by means of a toy exampleI proceed as follows. The "skeleton" of the market is given by a structure of a balance tree with branching and height equal to 3 and 4 respectively, as in figure B.1. Therefore, the network is composed of 40 stocks and 39 links. Starting from the center of that network, the links corresponding to the first, second, and third step away from it are given values 0.5, 0.3 and 0.15, respectively (see black letters in figure B.1). These numbers are going to take the role of the partial correlations between pair of the stocks. The weight for the rest of the links not considered in figure B.1 are assumed equal to zero.

With this structure at hand and the weighed adjacency matrix recovered using the expression (1), I compute the associated *real* covariance matrix.[11] Additionally, I assume a mean vector with components equal to 0.08 for the central node and 0.06, 0.04 and 0.02 for nodes located one, two, and three steps away from the center (see red figures in figure B.1). Using this covariance matrix and the commented mean vector as the *real* parameter of the process, I draw a sample of size 1.000 from a multivariate normal distribution. Next, I implement the 2-step estimation methodology explained at the beginning of this section assuming a confidence level of 1% upon this artificial dataset in order to estimate the *PCSN.*

Artificial partial correlation network FIGURE B.1



11 Since the matrix Δ only takes the role of normalization in equation (1), I consider as the relevant covariance just $-\rho^{-1}$ which is equivalent to saying that $\rho_{ij} = \beta_{ij}$.

Table B.1 provides a general assessment of the estimation performance by presenting the number of links correctly and incorrectly detected by the algorithm. Panel A shows the result of the estimation before filtering out the non-significant partial correlations. Panel B presents the same information after implementation of Monte Carlo Simulations.

Estimation algorithm's performance TABLE B.1

Initial Estimates (Panel A)

	Real Structure		
Estimated Structure	**Non Null**	**Null**	**Total**
Non Null	39	316	355
Null	0	425	425
Total	**39**	**741**	**780**

Final Estimates - Montecarlo Simulation (Panel B)

	Real Structure		
Estimated Structure	**Non Null**	**Null**	**Total**
Non Null	37	21	58
Null	2	720	722
Total	**39**	**741**	**780**

From panel A in table B.1, the first thing to notice is that the initial estimation does a good job of identifying the true network, although it is not perfect. Out of the 39 links that actually exist, all of them were initially captured. Its major problem is due to the number of links that the algorithm erroneously identified, say 316, leading to a network density of 45.5%. In general terms, the initial estimation provides good results in selecting the links that actually exist and discarding those not included in the real structure. The hit ratio of this estimation is (39+425)/780=59.5%.

Panel B from table B.1 shows the benefits of using Monte Carlo Simulation for uncovering the real network structure. The major change from panel A to panel B is due to the reduction of erroneously detected links. Such a reduction reaches 93%, from 316 to 21 leading to a higher hit ratio (37+720)/780=97.1%. Unfortunately, the algorithm discards two links that actually exist in the structure. Overall, the 2-step estimation methodology performs well and seems to be a suitable tool to uncover the hidden stock network.

Appendix C

Partial correlation matrix with optimal regularization

TABLE C.1

	ABS	AMS	ACE	BBVA	BKIA	BKT	POP	BSAB	SCH	BOLS	CTG	IDR	FERC	FCC	MAP	ANA	DIA	ENAG	CABK	OHL	GAM	IBE	IND	TL5	ACS	PROB	REE	REP	IAG	TECN	TEF	SCYR	VIS	JAZ	MITT
ABS																																			
AMS	–																																		
ACE	–	–																																	
BBVA	–	–	–																																
BKIA	–	–	–	–																															
BKT	–	–	–	0.1	0.1																														
POP	–	–	–	0.1	–	0.1																													
BSAB	–	–	–	–	0.1	0.1	0.3																												
SCH	–	–	–	0.6	–	–	–	–																											
BOLS	–	–	–	–	–	–	–	–	–																										
CTG	–0.1	–	–	–	–	–	–	–	–	–																									
IDR	0.1	0.1	–	–	–	–	–	–	–	–	–																								
FERC	–	–	0.1	–	–	–	–	–	–	–	–	–																							
FCC	–	–	–	–	–	–	–	–	–	–	–	–	–																						
MAP	–	–	0.1	–	–	–	0.1	–	–	–	–	0.1	0.1	–																					
ANA	–	–	–	–	–	–	–	0.1	–	–	–	–	–	–	–																				
DIA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.1																			
ENAG	–	–	–	–	–	–	–	–	–	–	0.3	–	–	–	–	–	–																		
CABK	–	–	–	–	–	0.2	0.2	–	0.2	–	–	–	–	–	0.2	–	–	–																	
OHL	0.1	–	0.1	–	–	–	–	–	–	–	0.1	–	0.1	0.1	–	0.1	–	–	–																
GAM	–	–	–	–	–	0.1	–	0.1	–	–	–	–	–	–	–	–	–	–	–	–															
IBE	–	–	–	–	–	–	–	–	0.1	0.1	0.1	–	–	–	–	–	–	0.1	–	–	–														
IND	–	–	0.1	–	–	–	–	–	0.1	–	–	–	–	–	–	–	0.1	–	–0.1	–	–	–													
TL5	–	–	–	–	–	0.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–												
ACS	0.1	–	–	–	–	–	–	–	–	–	–	–	0.2	–	–	0.1	–	–	–	0.1	–	–	–	–											
PROB	–	–	0.1	–	–	–	–	–	–	–	–	–	0.1	–	–	–	–	–	–	–	0.2	–	–	–	–										
REE	0.1	–	–	–	–	–	–	–	–	0.1	0.1	–	–	–	–	0.1	–	0.2	–	–	–	–	–	–	–	–									
REP	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.1	–	–	–	–	–	0.1	–	–	–	–								
IAG	–	0.1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.1	–	–	0.1	–	–	–	–							
TECN	–	0.1	–	–	–	–	–	–	–	0.1	–	0.1	–	0.1	–	–	–	–	–	0.1	–	–	–	–	–	–	–	0.1	–						
TEF	–	–	–	–	–	–	–	–	0.1	–	–	0.1	–	–	–	–	–	–	–	–	–	0.2	–	–	–	–	–	0.2	–	–					
SCYR	0.2	–	–	–	0.1	–	–	–	–	–	–	–	–	0.2	–	–	–	–	–	–	–	–	–	–	–	–	–	0.1	–	–	–				
VIS	–	0.1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–0.1			
JAZ	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–		
MITT	–	–	–	–	–	–	–	–	0.1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.1	0.1	–	–	–	–	

Direct correlation matrix

TABLE C.2

	ABS	AMS	ACE	BBVA	BKIA	BKT	POP	BSAB	SCH	BOLS	CTG	IDR	FERC	FCC	MAP	ANA	DIA	ENAG	CABK	OHL	GAM	IBE	IND	TL5	ACS	PROB	REE	REP	IAG	TECN	TEF	SCYR	VIS	JAZ	MITT
ABS																																			
AMS	0.34																																		
ACE	0.42	0.43																																	
BBVA	0.42	0.36	0.55																																
BKIA	0.42	0.37	0.46	0.55																															
BKT	0.43	0.30	0.51	0.60	0.54																														
POP	0.38	0.34	0.46	0.60	0.50	0.57																													
BSAB	0.43	0.35	0.52	0.61	0.54	0.60	0.64																												
SCH	0.40	0.35	0.55	0.77	0.55	0.59	0.58	0.61																											
BOLS	0.30	0.33	0.46	0.44	0.34	0.39	0.40	0.39	0.45																										
CTG	0.22	0.23	0.45	0.42	0.35	0.41	0.34	0.38	0.42	0.31																									
IDR	0.44	0.43	0.50	0.48	0.44	0.43	0.47	0.44	0.46	0.39	0.32																								
FERC	0.43	0.37	0.58	0.57	0.44	0.51	0.48	0.54	0.57	0.41	0.42	0.44																							
FCC	0.36	0.33	0.44	0.39	0.38	0.38	0.39	0.39	0.40	0.38	0.39	0.41	0.45																						
MAP	0.35	0.39	0.54	0.57	0.43	0.50	0.54	0.51	0.58	0.43	0.41	0.49	0.55	0.40																					
ANA	0.39	0.27	0.45	0.49	0.44	0.40	0.44	0.51	0.49	0.34	0.37	0.40	0.46	0.41	0.39																				
DIA	0.33	0.33	0.45	0.46	0.34	0.40	0.35	0.36	0.45	0.31	0.40	0.39	0.46	0.31	0.44	0.41																			
ENAG	0.19	0.21	0.34	0.28	0.26	0.27	0.20	0.28	0.27	0.26	0.50	0.28	0.33	0.29	0.30	0.30	0.27																		
CABK	0.38	0.28	0.48	0.59	0.52	0.58	0.58	0.57	0.61	0.41	0.37	0.39	0.47	0.42	0.55	0.42	0.39	0.23																	
OHL	0.44	0.38	0.55	0.50	0.42	0.46	0.47	0.53	0.50	0.41	0.46	0.44	0.55	0.46	0.49	0.51	0.41	0.33	0.48																
GAM	0.41	0.35	0.46	0.51	0.46	0.52	0.48	0.53	0.49	0.39	0.41	0.37	0.46	0.39	0.45	0.44	0.33	0.31	0.47	0.41															
IBE	0.32	0.33	0.55	0.58	0.43	0.51	0.43	0.52	0.60	0.47	0.52	0.42	0.55	0.40	0.51	0.49	0.45	0.42	0.47	0.48	0.45														
IND	0.32	0.34	0.49	0.52	0.37	0.39	0.36	0.36	0.52	0.33	0.40	0.39	0.48	0.33	0.45	0.34	0.42	0.26	0.33	0.42	0.37	0.45													
TL5	0.40	0.35	0.42	0.49	0.45	0.51	0.44	0.48	0.47	0.38	0.36	0.41	0.48	0.34	0.42	0.38	0.38	0.27	0.42	0.41	0.44	0.43	0.39												
ACS	0.45	0.39	0.57	0.59	0.50	0.52	0.47	0.54	0.59	0.42	0.48	0.49	0.60	0.43	0.52	0.51	0.45	0.38	0.47	0.56	0.48	0.57	0.44	0.47											
PROB	0.39	0.35	0.48	0.43	0.39	0.41	0.41	0.44	0.43	0.31	0.36	0.42	0.48	0.30	0.39	0.42	0.37	0.29	0.37	0.41	0.46	0.43	0.36	0.38	0.46										
REE	0.38	0.29	0.47	0.40	0.37	0.37	0.33	0.41	0.39	0.41	0.48	0.38	0.38	0.35	0.35	0.46	0.32	0.45	0.35	0.45	0.42	0.49	0.30	0.35	0.48	0.39									
REP	0.33	0.34	0.46	0.55	0.42	0.42	0.41	0.43	0.57	0.38	0.43	0.43	0.50	0.38	0.47	0.42	0.45	0.27	0.46	0.46	0.42	0.53	0.49	0.37	0.50	0.36	0.39								
IAG	0.39	0.39	0.47	0.48	0.39	0.45	0.41	0.43	0.45	0.34	0.31	0.41	0.46	0.35	0.39	0.39	0.37	0.25	0.42	0.40	0.46	0.41	0.34	0.44	0.46	0.39	0.33	0.33							
TECN	0.34	0.40	0.47	0.45	0.38	0.36	0.36	0.42	0.46	0.44	0.43	0.45	0.48	0.43	0.45	0.40	0.39	0.35	0.33	0.49	0.40	0.47	0.42	0.35	0.47	0.35	0.39	0.47	0.32						
TEF	0.37	0.37	0.55	0.63	0.47	0.50	0.48	0.52	0.64	0.44	0.45	0.50	0.54	0.43	0.54	0.45	0.46	0.30	0.49	0.51	0.41	0.61	0.47	0.44	0.57	0.40	0.38	0.58	0.45	0.47					
SCYR	0.47	0.39	0.45	0.53	0.51	0.48	0.51	0.54	0.52	0.39	0.42	0.46	0.51	0.47	0.46	0.46	0.40	0.26	0.46	0.49	0.46	0.47	0.38	0.39	0.51	0.43	0.43	0.48	0.41	0.45	0.47				
VIS	0.16	0.31	0.30	0.31	0.18	0.24	0.21	0.29	0.31	0.26	0.29	0.29	0.35	0.28	0.34	0.22	0.38	0.21	0.26	0.34	0.21	0.35	0.31	0.31	0.29	0.20	0.20	0.33	0.21	0.34	0.33	0.19			
JAZ	0.31	0.29	0.33	0.39	0.38	0.36	0.34	0.35	0.38	0.31	0.28	0.34	0.38	0.31	0.34	0.31	0.26	0.25	0.38	0.32	0.37	0.38	0.34	0.35	0.39	0.34	0.27	0.33	0.33	0.30	0.35	0.36	0.20		
MITT	0.31	0.31	0.43	0.47	0.39	0.37	0.41	0.42	0.49	0.33	0.30	0.35	0.41	0.36	0.42	0.35	0.33	0.19	0.41	0.38	0.36	0.38	0.36	0.34	0.40	0.36	0.26	0.35	0.39	0.42	0.43	0.37	0.26	0.30	

Appendix D

Centrality for each of the firms in IBEX-35 considering the most recent 250 trading days in the sample

Centrality by firm TABLE D.1

Ticker	NAME	Industry	Centrality
SCH	BANCO SANTANDER	Bank/Savings & Loan	0.593
BBVA	BBV.ARGENTARIA	Bank/Savings & Loan	0.557
CABK	CAIXABANK	Bank/Savings & Loan	0.249
POP	BANCO POPULAR ESPANOL	Bank/Savings & Loan	0.254
BKT	BANKINTER 'R'	Bank/Savings & Loan	0.227
BSAB	BANCO DE SABADELL	Bank/Savings & Loan	0.179
TEF	TELEFONICA	Utility	0.181
MAP	MAPFRE	Insurance	0.110
MITT	ARCELORMITTAL (MAD)	Industrial	0.111
IND	INDITEX	Industrial	0.062
BKIA	BANKIA	Bank/Savings & Loan	0.065
GAM	GAMESA CORPN.TEGC.	Industrial	0.073
REP	REPSOL YPF	Industrial	0.076
IBE	IBERDROLA	Utility	0.145
TL5	MEDIASET ESPANA COMUNICACION	Industrial	0.055
ACE	ABERTIS INFRAESTRUCTURAS	Industrial	0.041
IDR	INDRA SISTEMAS	Industrial	0.053
TECN	TECNICAS REUNIDAS	Industrial	0.055
FERC	FERROVIAL	Industrial	0.036
OHL	OBRASCON HUARTE LAIN	Industrial	0.042
ANA	ACCIONA	Industrial	0.047
SCYR	SACYR	Industrial	0.027
PROB	GRIFOLS ORD CL A	Industrial	0.026
AMS	AMADEUS IT HOLDING	Industrial	0.024
IAG	INTL.CONS.AIRL.GP. (MAD) (CDI)	Transportation	0.038
CTG	GAS NATURAL SDG	Utility	0.059
ENAG	ENAGAS	Utility	0.055
ABS	ABENGOA B SHARES	Industrial	0.017
REE	RED ELECTRICA CORPN.	Utility	0.041
ACS	ACS ACTIV.CONSTR.Y SERV.	Industrial	0.021
BOLS	BOLSAS Y MERCADOS ESPANOLES	Other Financial	0.037
FCC	FOMENTO CONSTR.Y CNTR.	Industrial	0.017
DIA	DISTRIBUIDORA INTNAC. DE ALIMENTACION	Industrial	0.022
VIS	VISCOFAN	Industrial	0.006
JAZ	JAZZTEL	Industrial	0.000

Appendix E

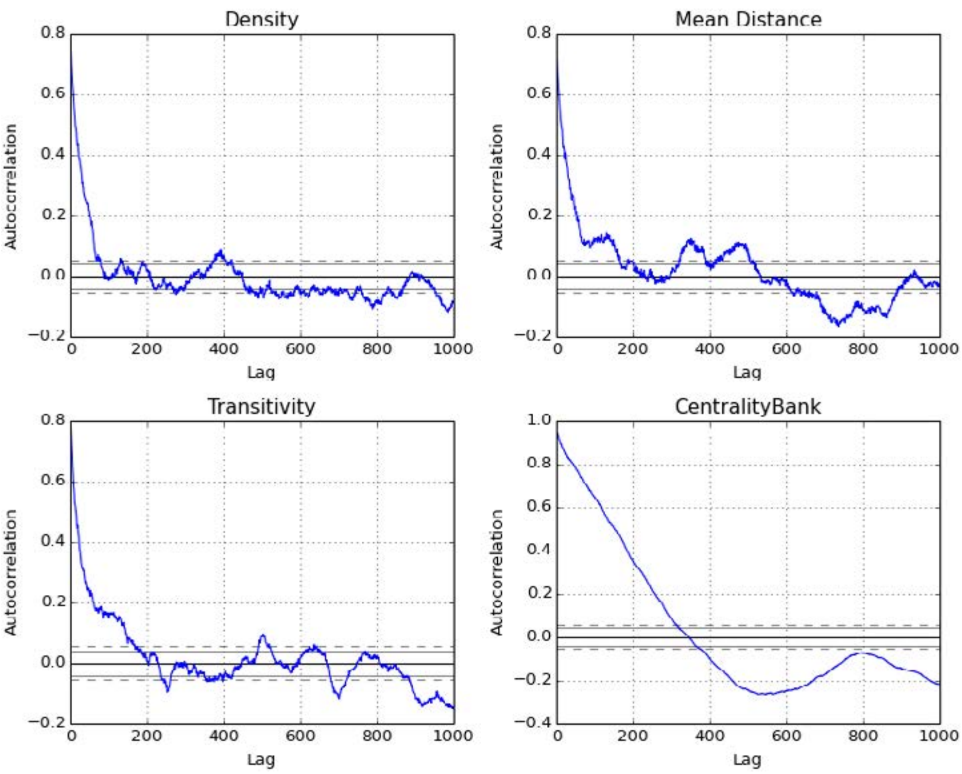
Correlation between regressors

TABLE E.1

	Density	Mean Distance	Transitivity	CentralityBank
Density				
Mean Distance	-0.702			
Transitivity	0.178	0.210		
CentralityBank	0.153	0.147	0.368	

Autocorrelation function of regressors

FIGURE E.1



Appendix F

Mean and total centrality by economic sectors - France TABLE F.1

		Eigenvector Centrality		Market Value*	
France	**Number of Firms**	**Total**	**Mean**	**Total**	**Mean**
Other Financial	1	0.049	0.049	19,969.1	19,969.1
Transportation					
Insurance	1	0.335	0.335	47,251.2	47,251.2
Utility	3	0.071	0.024	127,758.0	42,586.0
Industrial	32	2.413	0.075	863,541.5	26,985.7
Bank/Savings & Loan	3	1.302	0.434	128,771.3	42,923.8
Total	**40**	**4.171**	**0.104**	**1,187,291.0**	

* In Millons of €.

Mean and total centrality by economic sectors - Germany TABLE F.2

		Eigenvector Centrality		Market Value*	
Germany	**Number of Firms**	**Total**	**Mean**	**Total**	**Mean**
Other Financial					
Transportation	2	0.162	0.081	36,468.0	18,234.0
Insurance	2	0.468	0.234	85,546.1	42,773.0
Utility	3	0.915	0.305	100,876.6	33,625.5
Industrial	21	2.487	0.118	654,366.7	31,160.3
Bank/Savings & Loan	2	0.327	0.163	51,514.2	25,757.1
Total	**30**	**4.359**	**0.145**	**928,771.7**	

* In Millons of €.

Mean and total centrality by economic sectors - Italy TABLE F.3

		Eigenvector Centrality		Market Value*	
Italy	**Number of Firms**	**Total**	**Mean**	**Total**	**Mean**
Other Financial	2	0.243	0.122	7,011.4	3,505.7
Transportation					
Insurance	2	0.210	0.105	31,034.9	15,517.4
Utility	6	0.399	0.067	87,142.5	14,523.7
Industrial	20	0.543	0.027	194,897.5	9,744.9
Bank/Savings & Loan	7	2.266	0.324	95,289.8	13,612.8
Total	**37**	**3.662**	**0.099**	**415,376.0**	

* In Millons of €.

Mean and total centrality by economic sectors - UK

TABLE F.4

		Eigenvector Centrality		Market Value*	
UK	**Number of Firms**	**Total**	**Mean**	**Total**	**Mean**
Other Financial	8	0.0033	0.0004	44,977.0	5,622.1
Transportation	3	0.0013	0.0004	17,665.3	5,888.4
Insurance	10	0.0371	0.0037	104,199.3	10,419.9
Utility	7	0.0003	0.0000	159,222.0	22,746.0
Industrial	68	2.6636	0.0392	1,219,609.4	17,935.4
Bank/Savings & Loan	5	0.0012	0.0002	263,657.1	52,731.4
Total	**101**	**2.7068**	**0.027**	**1,809,330.2**	

* In Millons of £.