A Network Approach to Portfolio Selection

Gustavo Peralta\textsuperscript{a}, Abalfazl Zareei\textsuperscript{b}

This Version: November 13, 2014

Abstract

A financial market can be conceived as a network with stocks as nodes and links accounting for return’s correlation. We prove, theoretically, the relation between the centrality of each individual stock in the network structure and the weight it receives in an optimal allocation of wealth. Therefore and consistent with the diversification rule, optimal portfolios assign wealth toward the periphery of the market network in order to reduce the influence of central assets. Next, we empirically investigate the major determinants for stock’s centrality and its relative stability. We show that both, financial and market variables are fundamental drivers of stock’s centrality. Our results also make evident the increased stability of those stocks that are centrally placed in the network. Finally, we explore by means of in-sample and out-of-sample analysis, how network-based investment strategies could be used to enhance portfolio’s performance. The major contribution of the paper regards to a simplification in the portfolio selection process via targeting a group of stocks belonging to certain region of the stock market network.

Keywords: Network Theory, Network Centrality, Portfolio Selection

JEL Classification: C00, C45, C55, G10, G11, G17

\textsuperscript{a} Department of Research and Statistics at CNMV and Department of Business Administration at Universidad Carlos III de Madrid. Email: gperalta@cnmv.es

\textsuperscript{b} Department of Business Administration at Universidad Carlos III. Email: azareei@emp.uc3m.es.
1. Introduction

In his seminal paper, (Markowitz 1952) laid the foundation for the modern theory of portfolio selection in which the covariance matrix of stock returns is a fundamental ingredient in the optimal portfolio’s recipe. Despite the profound changes that this publication led in the financial literature, the out-of-sample performance coming out from Markowitz’s prescriptions are not as promising as it was expected (Jobson & Korkie 1980), (Michaud 2008), (DeMiguel et al. 2009) and (Duchin & Levy 2009). Recently, several researchers have used the correlation (covariance) of returns to characterize the stock markets as a complex network ([Mantegna 1999], (Vandewalle et al. 2001), (Bonanno et al. 2004), (Onnela et al. 2003b), (Tse et al. 2010)). In the last few years, the financial literature became interested in the network approach to shed some light on systemic risk issues in such markets (Billio et al. 2012), (Diebold & Yilmaz 2014), (Hautsch et al. 2014). This paper embraces this innovative framework to investigate the extent to which the knowledge about the topology of such a network could be used as an effective tool in the portfolio selection process.

We characterize each stock in the market based on two major attributes: its individual performance and its systemic performance. The individual performance regards to its performance in isolation of the rest of the firms. It could be accurately measured by its standard deviation of return or by its Sharpe ratio (depending on the particular investment’s purpose). The systemic performance accounts for the role of a particular stock as a part of the whole market. It could be properly measured by the centrality scores of each firm into the stock market network. Throughout this paper, we delve into the interaction of these two dimensions and their impacts on the optimal investment choices. Our major contribution regards to the reduction in complexity implied in the portfolio selection process by targeting group of stocks within a specified range of centrality, thus focusing in a subset of firms belonging to certain regions of the stock market network. Building upon this network-based strategy, we show that higher out-of-sample risk-adjusted returns could be obtained compared to others well-established strategies.

The paper rests upon three major sections. In the first part we show, theoretically, the negative relationship between the optimal portfolio’s weights and stock’s centrality. Therefore, the dispersion of the optimal amount of weight assigned to each asset in a portfolio mirrors the heterogeneity of their centrality. Then, in our framework, Markowitz’s rule means that the amount of weight assigned to extremely central firms in the network should be lower compared to less central ones. The intuition behind this result is straightforward; highly central stocks in a correlation-based network lead the general movement in the market and thus reduce portfolio’s performances by increasing its variance.

In the second part we attempt to answer two related questions: i) what are the major drivers of stock’s centrality and ii) to which extent there is a tendency for highly central firms to remain in the market, a feature that we called stability. To be more precise, in i) we empirically investigate the major determinants of stock’s centrality. We find that both, financial and market variables affect stock’s centrality, however the evidence is stronger for the second group. In particular, our results show that leverage, market capitalization and stock return positively effects stock’s centrality, but
liquidity, traded volume, market prices and the standard deviation of return show a negative impact. By including into the analysis the full list of firms that have ever existed in NYSE in a particular period of time, in ii) we study the relative stability of stocks in accordance to their centrality. The results show that the percentage of low-central stocks being delisted is significantly higher than both, medium-central and high-central stocks. Further, for those firms that remains listed, we detect a clear tendency to maintain its centrality status throughout large period of the time.

Finally, in the third section, we test the extent to which the topological structure of the market network could be used as an effective investment tool. We do this by means of in-sample and out-of-sample approaches. Considering the most interesting case of out-of-sample analysis, we rely on three different datasets (two of them for the US market with daily and monthly data and a third one with daily data for the UK market) to check the consistency of our results. We find that the association between the systemic performances and the individual performances of stocks is time and market dependent. Therefore, when peripheral stocks show higher individual performances relative to central ones, no trade-off exists and a larger amount of wealth should be allocated toward the outskirt of the market network. However, when the stock’s individual performances are positively associated with their centrality, a convenient investment should balance these two confronting forces. Defining \( \rho \) as the correlation between stock’s systemic performances (centralities) and stock’s individual performances (Sharpe ratios), we compute out-of-sample risk adjusted returns coming out from what we call \( \rho \)-dependet strategy. Such strategy works as follows. If \( \rho \) is not sufficiently high, naïve strategy (consisting in weights equal to \( 1/n \) for \( n \) as the number of stocks) is applied upon a set of peripheral stocks. As long as \( \rho \) assumes values larger than a predefined threshold, \( \tilde{\rho} \), a naïve strategy is applied toward a set of stocks belonging to the center of the market network. Our results show better out-of-sample performance of this \( \rho \)-dependet strategy when they are compared to well-known benchmarks. We recognize that the major drawback of the approach regards on how to define \( \tilde{\rho} \).

To the extent of our knowledge, (Pozzi et al. 2013) is the only paper that used the topology of the market network for investment purposes. They show the convenience to invest in stocks belonging only to the periphery of the market network. Our major conclusions are not perfectly in line with theirs but neither in a total disagreement. In their study, the relationship between systemic performance and individual performance is missed leading to an unconditional investment strategy focus only in the outskirt of the market. Different from this, our investment strategy is somehow more organic and contingent to the state of the market throughout the use of \( \rho \) in its implementation. Additionally, in (Pozzi et al. 2013), the authors do not justify the use of any network centrality measure, instead they combine several of them into a synthetic centrality index. In the current paper, the centrality measures used through the analysis stems from the theoretical framework connecting network centrality with optimal portfolio weights.

---

1 The in-sample approach is only considered in order to get some insights about the way in which network measures affect the optimal portfolio choices.
The remainder of the paper is organized as follows. Section 2 reviews theoretically the definition of node’s centrality used in this study. This section also embeds this measure of centrality into Markowitz’s framework for portfolio selection. Section 3 describes the main statistical features of stocks in accordance to their centrality. We estimate a panel regression model in order to explain the major financial and market determinant for the centrality of each stock. In section 4, we present results regarding the relative stability of different assets in terms of their centrality profiles. Section 5 describes empirically the in-sample relationship between centrality and optimal portfolio’s weights with a strong emphasis on the interaction between the individual and systemic stock’s performances. In section 6, we analyze the extent to which the structure of the market network could be used as an effective tool in enhancing out-of-sample portfolio’s performance. Finally, section 7 concludes and outlines future research lines.

2. Optimal Portfolio Weights and Network Centrality

There is a large debate in the network-related literature, particularly in sociology regarding how to measure the centrality of a particular agent who is embedded into a network of relationships. The importance of such measures stems from the implicit assumption of the incremental power/status attached to highly central individuals. Despite its intuitive meaning, the whole concept of centrality is somehow ambiguous and its measurements depend on the particular underlying process. For example, in a social network, agent \( i \) becomes central when he/she interacts with other central agents. Contrary to this, in a bargaining process, the centrality of agent \( i \) stems from its connection with other non-central agents. In his seminal papers (Bonacich 1972), the author proposes a measure of centrality that have become the standard in the network literature, the so called **eigenvector centrality**. The next subsections formalize such a notion in a capital market context and determine its relationship with the optimal portfolio’s weights.

2.1 Measuring Centrality

In general terms, a network is a pair of sets \( G = \{N, \Omega\} \), with \( N = \{1,2,...,n\} \) as the set of nodes and \( \Omega \) is the set of links connecting a pair of them. Then, if there is a link from node \( i \) to node \( j \), \((i,j) \in \Omega\). A convenient way to arrange the information contained in \( \Omega \) is by means of the so-called adjacency matrix, \( \Omega = [\Omega_{ij}] \). \( \Omega \) is an \( nxn \) matrix in which \( \Omega_{ij} \neq 0 \) captures the existence of a relationship between node \( i \) and \( j \). The network is said to be **undirected** if \( \Omega = \Omega^T \), therefore if \((i,j) \in \Omega\) automatically implies \((j,i) \in \Omega\). Note that for undirected network, no causal relationship is attached to the links and they are visually represented as a line, \((j \rightarrow i)\). On the other hand, if \( \Omega \neq \Omega^T \), the network is said to be **directed** and \( \Omega_{ij} \) entails a causal relationship from node \( j \) to node \( i \) which does not necessarily implies the reverse. In this case, the links are visually represented as arrows, \((j \rightarrow i)\). Further, if \( \Omega_{ij} \in \{0,1\} \), \( G \) is said to be **un-weighted**. However, when \( \Omega_{ij} \in \mathbb{R} \), links carry 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.

As it is stated in (Bonacich 1972), eigenvector centrality assumes that the centrality of node $i$, $v_i$, is proportional to the weighted sum of its neighbor’s centrality:

$$ v_i \equiv \lambda^{-1} \sum_j \Omega_{ij} v_j $$

(1)

By restating equation (1) in matrix terms, the eigenvector centrality, $v$, is given by the eigenvector of $\Omega$ corresponding to the eigenvalue $\lambda$, where the largest eigenvalue is the preferred choice.\(^2\)

$$ \lambda v = \Omega v $$

(2)

**Definition 1**: Consider an undirected and weighted network $G = \{N, \Omega\}$, with $N$ as the set of nodes and $\Omega$ as its adjacency matrix. Then, the eigenvector centrality of node $i$ is proportional to the $i$-th component of the eigenvector of $\Omega$ corresponding to the largest eigenvalue, $\lambda_1$. The factor of proportionality is $\lambda_1^{-1}$

Note that equation (1) shows that a node becomes central either by being connected with many other nodes (with positive centrality) or by being connected with just few highly central ones. This quantity is suitable to be computed for weighed and un-weighed networks as well. However, for the directed structure, such centrality measure presents shortcomings that make it not advisable for its implementation.

### 2.2 Main result from portfolio’s selection theory

In (Markowitz 1952), the pillars for the modern portfolio theory were established. Next, we briefly summarize his fundamental results as we are going to heavily rely on them. Let us assume $n$ risky assets with a vector of expected return denoted as $\mu$ and covariance matrix as $\Sigma$. Consider the problem of finding the vector of optimal weights, $w$, that minimizes the variance of such portfolio subjected to $w^T 1 = 1$. This strategy is commonly known as minimum-variance strategy or $m$-var for short. Formally:

$$ \text{Min} \sigma_p^2 = w^T \Sigma w \quad \text{subject to} \quad w^T 1 = 1 $$

(3)

The solution of (3) is given by:

\(^2\) In principle, all eigenvector of $\Omega$ is a solution for in (2). However, the score of centralities corresponding to the largest component in the network is given by the largest eigenvalue (Bonacich 1972)
\[ w_{m}^{*} = \frac{1}{1\Sigma^{-1}\Sigma^{-1}1} \]

Let us define the correlation matrix of return as \( \Omega \), the standard deviation of return for stock \( i \) as \( \sigma_i \) and \( \Delta = \text{diag}(\sigma) \). Then, the connection between the correlation and the covariance matrices is given by \( \Sigma = \Delta \Omega \Delta \). Restating expressions (4) in terms of \( \Omega \) gives

\[ \hat{w}_{m}^{*} = \varphi_{m} \Omega^{-1} \epsilon \]

where \( \hat{w}_{i,m}^{*} = w_{i,m}^{*} \sigma_i, \varphi_{m} = \frac{1}{1\Sigma^{-1}1} \) and \( \epsilon_i = 1/\sigma_i \).

Considering a closely related problem to (3) which includes a riskless asset with return \( r_f \). Thus, the portfolio is composed of \( n + 1 \) assets, \( n \) risky assets and a risk-free asset. We denote the excess return of asset \( i \) \( (r_i - r_f) \) as \( \hat{r}_i \) and the vector of expected excess return as \( \hat{\mu}_e \). The problem of minimizing the variance of the portfolio for a given level of excess return \( R_e \) is expressed as

\[ \min \sigma_p^2 = w^T \Sigma w \quad \text{subject to} \quad w^T \hat{\mu}_e = R_e \]  

(6)

The strategy implied by (6) is known as the mean-variance strategy or \( M\text{-var} \) for short. Note that \( w^T 1 = 1 \) is not a restriction in (6) since part of the investor’s wealth could be allocated to the risk-free asset, \( w_f = 1 - w^T 1 \). Nevertheless, when the tangency portfolio is considered, \( w_f = 0 \). In any case, the optimal solution for the \( M\text{-var} \) strategy is given by

\[ w^* = \frac{R_e}{\hat{\mu}_e^T \Sigma^{-1} \hat{\mu}_e} \Sigma^{-1} \hat{\mu}_e \]  

(7)

By following the same logic as before, expression (7) could be restated in terms of the correlation matrix as follows

\[ \hat{w}^* = \varphi \Omega^{-1} \hat{\mu}_e \]

(8)

where \( \hat{w}_{i}^{*} = w_{i}^{*} \sigma_i, \varphi = \frac{R_e}{\hat{\mu}_e^T \Sigma^{-1} \hat{\mu}_e} \) and \( \hat{\mu}_i^e = \mu_i^e / \sigma_i \).

2.3 The relationship between Optimal Portfolio’s Weights and Stock Centralities

Let us assume a capital market network \( CM = \{N, \Omega \} \) with \( N \) as the set of stocks and \( \Omega \) as the adjacency matrix given by the correlation matrix of returns. We set the main diagonal of \( \Omega \) to zero in order to discard meaningless self-loops from the structure. As it is proved in appendix A, this does not change neither the structure of its eigenvectors nor the order of its eigenvalues. Next,
Proposition 1 and Corollary 1 state the relationship between the optimal weight of asset $i$ in terms of its eigenvector centrality under the \textit{m-var} and \textit{M-var} strategies. The reader is directed to appendix \textit{A} for the corresponding proof.

\textbf{Proposition 1:} Consider a capital market network $CM = \{N, \Omega\}$ and $\{\nu_1, ... \nu_n\}$ and $\{\lambda_1, ... \lambda_n\}$ the sets of eigenvectors and eigenvalues (in descendent order) of $\Omega$, respectively. Then, the optimal portfolios weights in (5) and (8) could be written as:

\[ \hat{w}^*_{mv} = \varphi_{mv} \epsilon + \varphi_{mv} \left( \frac{1}{\lambda_1} - 1 \right) \epsilon_M \nu_1 + \Gamma_{mv} \]  
\[ \hat{w}^* = \varphi \hat{\mu}_e + \varphi \left( \frac{1}{\lambda_1} - 1 \right) \hat{\mu}_M^e \nu_1 + \Gamma \]  

Where $\epsilon_M = (\nu_1^T \epsilon)$ and $\hat{\mu}_M^e = \nu_1^T \hat{\mu}_e$

Proposition 1 clearly determines the connection between optimal weights and the first eigenvector of the correlation matrix which in network theory represents eigenvector centrality. Additionally, from the perspective of Principal Component analysis, we can interpret $\epsilon_M$ and $\hat{\mu}_M^e$ as the inverse of the standard deviation of returns and the return-to-risk ratio \textit{at market level}. The first term in expressions (9) and (10) captures the naïve idea of investing in an asset taking into account only its risk or its return-to-risk ratio, respectively, as it were in isolation. Then, the lower (higher) the standard deviation of returns (return-to-risk ratio) of asset $i$, the higher must be its optimal portfolio’s weight. We call this as the \textit{individual performance} of asset $i$. The second terms in expression (9) and (10) account for the centrality of any given asset in the capital market network. We call this as the \textit{systemic performance} of asset $i$. Corollary 1 establishes a negative relationship between the amount of wealth devoted to a particular asset and its eigenvector centrality.

\textbf{Corollary 1:} Assume that $\lambda_1 > 1$ and that $\epsilon_M$ and $\hat{\mu}_M^e$ are positive number. Then, a large eigenvector centrality of stock $i$ is consistent with a low optimal portfolio’s weight in both, \textit{m-var} or \textit{M-var} strategy.

Corollary 1 states that, under plausible conditions, those stocks that receive large amount of wealth in an optimal portfolio tend to be placed toward the outskirt of the network. This statement is consistent with the results in (Pozzi et al. 2013). However, in that paper the individual performances are totally discarded from the analysis as well as the interaction with the systemic performance (centrality). In section 5 and 6, we present evidence showing that this association is
time and market dependent. Therefore, there are periods of time in which the most systemic (central) stocks are also the best individually performing assets leading to a trade-off in the investment portfolio choice.

3. What determines Stock Centrality?

This section is devoted to investigate the major financial and market drivers of stock’s centrality and to characterize the salient features of stocks in terms of their centrality profile. Our period of analysis starts from October 2002 until December 2012. We rely on CRSP data base for daily information about split and dividends-adjusted stock prices and returns. We merged this information with quarterly COMPUSTAT files in order to obtain financial data for the selected companies. The firms included into the sample correspond to the 200 most capitalized components of the S&P 500 in 2013 discarding: i) those companies with negative total equity in any of the quarter under analysis and ii) those companies with missing values on market prices or returns. Throughout the paper we call this data set as d-S&P500. Our final sample consists of 200 stocks over 2,581 trading days and 7,931 firms-quarter data points. For a complete list and description of every datasets used in this paper, see appendix D.

We calculate the shrinking covariance (correlation) matrix of returns following (Ledoit & Wolf 2004) using the information about excess returns. We denote these robust estimators for the covariance and correlation matrices of returns as $\hat{\Sigma}$ and $\hat{\Omega}$, respectively. It worth mentioning that when $\hat{\Omega}$ takes the place of the adjacency matrix of a market network, its main diagonal is set to zero to prevent meaningless self-loops. For the rest of the paper, we will refer to centrality and eigenvector centrality as synonymous.

Let us start our empirical analysis by presenting the main descriptive statistics about stock centralities. Table 1 splits the sample of firms into different industrial sectors (first 2 digit from the SIC codes) showing their total and mean centrality. Note that regardless that most of the traded volume and market capitalization comes from the manufacturing sector, the maximum centrality per company corresponds to those belonging to the financial sectors. An observation which is consistent with (Tse et al. 2010).

---

3 The corresponding monthly risk-free rate is taken from Kenneth French’s Website.
In figure 1, box plots for Sharpe ratio and excess return are plotted for different levels of centrality. Three exclusive and collectively exhaustive groups of stocks are created: highly, medium and low central nodes belong to the bottom, middle and top 33.3% quantiles of the centrality distribution (see appendix B for a complete tablet with descriptive statistics). Despite that there is not a significant difference in mean or median, the distribution of each of the performance variables tends to become more concentrated as long as higher centrality scores are taken into account. Therefore, more heterogeneous performances are allocated to the periphery of the market network.

In order to understand how centrality ranks stocks according to their average correlation with rest of the markets, we present the left panel of figure 2 where a clear positive and non-linear relationship is evident. Additionally, the right panel from the same figure shows the scatter plot between the stocks’ centrality and its corresponding Beta-CAPM assuming the S&P 500 as the market index. As it is expected, eigenvector centrality is positively associated with Beta CAPM but the dispersion around this relationship suggests that the first variable also captures other type of features in the data.

Table 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Agriculture, Forestry, And Fishing</td>
<td>0</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>B</td>
<td>Mining</td>
<td>17</td>
<td>9%</td>
<td>656,688</td>
<td>6%</td>
<td>26,178</td>
<td>7%</td>
</tr>
<tr>
<td>C</td>
<td>Construction</td>
<td>2</td>
<td>1%</td>
<td>44,905</td>
<td>0%</td>
<td>3,073</td>
<td>1%</td>
</tr>
<tr>
<td>D</td>
<td>Manufacturing</td>
<td>87</td>
<td>44%</td>
<td>5,338,584</td>
<td>48%</td>
<td>156,881</td>
<td>43%</td>
</tr>
<tr>
<td>E</td>
<td>Transportation, Communications, Electric, Gas, And Sanitary Services</td>
<td>27</td>
<td>14%</td>
<td>1,145,092</td>
<td>10%</td>
<td>36,988</td>
<td>10%</td>
</tr>
<tr>
<td>F</td>
<td>Wholesale Trade</td>
<td>5</td>
<td>3%</td>
<td>142,423</td>
<td>1%</td>
<td>6,011</td>
<td>2%</td>
</tr>
<tr>
<td>G</td>
<td>Retail Trade</td>
<td>12</td>
<td>6%</td>
<td>745,486</td>
<td>7%</td>
<td>27,130</td>
<td>7%</td>
</tr>
<tr>
<td>H</td>
<td>Finance, Insurance, And Real Estate</td>
<td>37</td>
<td>19%</td>
<td>2,254,158</td>
<td>20%</td>
<td>70,788</td>
<td>19%</td>
</tr>
<tr>
<td>I</td>
<td>Services</td>
<td>13</td>
<td>7%</td>
<td>829,671</td>
<td>7%</td>
<td>37,314</td>
<td>10%</td>
</tr>
<tr>
<td>J</td>
<td>Public Administration</td>
<td>0</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
<td>-</td>
<td>0%</td>
</tr>
<tr>
<td>Total</td>
<td>200</td>
<td>11,157,008</td>
<td>364,364</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Box plot of individual performance measures for different level of stock’s centrality.

In order to understand how centrality ranks stocks according to their average correlation with rest of the markets, we present the left panel of figure 2 where a clear positive and non-linear relationship is evident. Additionally, the right panel from the same figure shows the scatter plot between the stocks’ centrality and its corresponding Beta-CAPM assuming the S&P 500 as the market index. As it is expected, eigenvector centrality is positively associated with Beta CAPM but the dispersion around this relationship suggests that the first variable also captures other type of features in the data.
Next, we move on toward understanding the major drivers of stock’s centrality. More specifically, we investigate whether financial and market variables have significant impact on the stocks centrality. To do this, the next quarterly-based panel regression is estimated:

\[
Centrality_{it} = \beta_0 + \beta_1 Profitability_{it} + \beta_2 Leverage_{it} \\
+ \beta_3 Liquidity_{it} + \beta_4 M\_to\_B_{it} + \beta_5 \ln(MV_{it}) + \beta_6 \ln(TV_{it}) + \beta_7 RetQ_{it} \\
+ \beta_8 \ln(Px_{it}) + \beta_9 StdRet_{it} + \epsilon_{it}
\]

(11)

where the dependent variable, \(Centrality_{it}\), stands for the eigenvector centrality of firm \(i\) in period \(t\). For selecting relevant financial and market variables, we rely on (Campbell et al. 2008) to identify the set of measures that has been proved to affect firms under distress. Among the first group, we include: \(Profitability\) which is calculated as the ratio of net income to total assets, \(Leverage\) computed as total liability relative to total assets and finally, \(Liquidity\) calculated as the ratio between cash and short term asset to total assets. Among the second group of explanatory variables, we include \(\ln(MV)\) accounting for the logarithm of the firm’s market value (market price multiplied by total common stocks), \(\ln(TV)\) which regards to the logarithm of total trades in the particular quarter, \(RetQ\) standing for the quarterly excess return, \(\ln(Px)\) which accounts for the logarithm of the market price at the end of the respective quarter and finally \(StdRet\) which considered the daily standard deviation of returns for the corresponding quarter. Finally, \(M\_to\_B\) stands for Market-to-Book ratio calculated as the ratio between firm’s market value to firm’s book value of common stocks. To control for outlier we windsorize each measure at 1%. Table 2 reports summary statistics for all the commented variables. As a final comment, we match the centrality of each stock at the last day of a given quarter with the financial ratios corresponding to that exact quarter.
Table 3 presents (in columns) the estimation of equation (11) for three different specifications. Columns I and II includes industry and quarter fixed effect in which the first one only considered robust-heteroscedastic standard errors while the second one introduces the correction in the standard errors by applying two-ways clustering (Petersen 2009). Column III introduces industry-dummy variables and clustering regression errors by quarters.

The results evidence a positive and significant effect for market value and leverage over stock’s centrality. Particularly interesting is the case of market values given its extremely low p-values. Contrary to this, ours calculations indicate that liquidity, traded volume and market prices affect negatively and significantly the centrality of a stock. For the case of profitability and market-to-book ratio, the regressions could not capture a significant effect in any of the specifications. For cases of quarterly market return and the daily standard deviation of returns, their coefficients are statistically significant, positive and negative respectively, only for the first specification. However, their significances are undermined in models II and III. It worth mentioning that since financial firms are shown to be the most central ones (see table 1), the patter of significant p-values seems to coincide somehow with the salient features of such sector (large market cap and leverage). We did not investigate further this characterization of economic sectors and its relationship with centrality leaving it as a line for future research.

In summary, we can identify highly central stocks as being characterized as highly leveraged and poorly liquid firms in terms of financial ratios. In terms of market variables, they show large market capitalization, a not so intense trading activity with market prices impacting negatively in its centrality.
### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Models I</th>
<th>Models II</th>
<th>Models III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Profitability</td>
<td>Leverage</td>
<td>Liquidity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.00663 (-0.88)</td>
<td>0.00220 (2.67)**</td>
<td>-0.00578 (-5.56)***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Market-to-Book</td>
<td>Market Value</td>
<td>Traded Volume</td>
</tr>
<tr>
<td></td>
<td>-0.0000427 (-0.77)</td>
<td>0.00275 (12.08)***</td>
<td>-0.00249 (-10.31)***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Quarterly Return</td>
<td>Price</td>
<td>Std of Return</td>
</tr>
<tr>
<td></td>
<td>0.00383 (4.47)***</td>
<td>-0.00271 (-7.94)***</td>
<td>-0.0538 (-3.09)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7931</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R^2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.171</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dummies</td>
<td></td>
<td>Coefficient estimates</td>
</tr>
<tr>
<td></td>
<td>Ind-Quarter</td>
<td>Ind-Quarter</td>
<td>Ind</td>
</tr>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>OLS</td>
</tr>
<tr>
<td></td>
<td>White</td>
<td>CL-Firm &amp; Quarter</td>
<td>CL-Quarter</td>
</tr>
</tbody>
</table>

* p<0.05, ** p<0.01, *** p<0.001

### 4. The relationship between Stocks Centrality and its Stability

We associate the concept of stock’s stability to the tendency of a particular asset to remain listed in the market through time without any change in its relative centrality status. This section analyzes the stability of stocks listed in NYSE from two different perspectives. First, we investigate the average percentage of stocks that are delisted from the market with respect to their centrality scores. Second, we present results regarding to the switching nature of assets in accordance to their different positions into the market network.
The dataset used in this section contemplates all of the NYSE stocks with monthly pricing records in the period starting from April-1968 until April-2012. Thus, we include a full list of companies that have ever existed at some point in time. Delisting firms are located in accordance to their presence in the data set. The specific delisting period for a particular company corresponds to the month in which the dataset provider stops reporting prices for that company. We called this dataset as m-NYSE and we rely on CRSP for its construction (see appendix D).

As it was commented, the first experiment investigates the propensity of stocks to remains in the market in accordance to their centrality. We select a 20 year moving window of returns in order to compute the centrality distribution for the full list of active stocks at that period. Then, three groups of stocks are created: highly, medium and low central in accordance to the top, middle and bottom 5% of the centrality distribution. Afterwards, we compute the corresponding percentage of those stocks that remain in the NYSE for each of the next 20 years. Since a one-month displacement step is considered for the moving window, we end up with 50 iterations of this process. Figure 3 shows the average of the percentages of stock remaining in the market over the 50 iterations for year 1 to 20.

![Figure 3: Average percentage of Stocks remaining listed in NYSE during the next 20 years in accordance to their centrality](image)

As it can be inferred from the above figure, the percentage of low central stocks that stay listed in NYSE is consistently lower throughout the 20 years under analysis compared to both, medium and high central stocks. We arrive to the same conclusion by confronting medium central stocks to the high central ones. Further, as long as we increase the distance away from period zero, these differences tends to increase. To be more concrete, one year apart from period zero, the average percentages of stocks that remain in the market from the high and low central groups correspond to

---

4 We track the delisting in CRSP according to the PERMNO of stocks. This number represents the entire trading history of stocks in CRSP despite changes in name or capital structure. We consider delisting of a stock to be the deletion of its PERMNO from CRSP time series dataset.
99.5% and 94.4%, respectively. However, those percentages for the 20\textsuperscript{th} year attain values of 40.4% and 23.2%, respectively.

In our second analysis, we turn to analyze the change in the nature of stocks in terms of centrality by relying also on a moving window approach. Therefore, we concentrate only in those stocks that remain in the market. We specify a 30-year moving window and divide it into two sub-periods, each of them with 15 years. We use the first and the second period to have an initial and final categorization of stocks in accordance to their centrality. Three exclusive and collectively exhaustive groups of stocks are created: highly, medium and low central in accordance to the top, middle and bottom 33.3% quintiles of the centrality distribution. Then, we construct a switching matrix accounting for the distribution of stocks belonging to a particular range of centrality in the initial period in terms of their centrality in the final period. Since one-year displacement step is under consideration, 15 individual switching matrices were computed. Figure 4 reports the results for each of those iterations\textsuperscript{5}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4}
\caption{Change in nature of stocks from an initial percentile of centrality to a final percentile of centrality across 15 iterations years}
\end{figure}

In the left-panel of figure 4, it is well presented the fact that most of low central stocks tend to change their nature to medium central or stay low central. Additionally, we can see that just a small proportion of low central stocks change to become highly central. The middle-panel of figure 4 depicts the changing nature for medium central stocks. As it can be inferred, medium central stocks mostly stay in the middle range of centrality and lower percentages of them tend to change toward low or high centrality. Finally, the right-panel of figure 4 presents the result of the experiment when the initial stocks were classified as highly central assets. We can observe that high central stocks tend to stay central across the iterations and just lower percentage of them tend to become low or medium central. As a sort of summary, figure 5 presents the average of the percentage of switching in centrality over the 15 interactions.

\footnote{The sum of horizontal height for each line does not sum up to one since the proportion of the delisted firms is not included in the graph.}

14
Two main conclusions should be highlighted from this section. First, low central stocks show a higher probability of being delisted from the market compared to the rest of the category. Second, among the stocks that remain in the market, there is a tendency to occupy the same position into the market network through time providing a sort of stability to the structure in terms of their nature of centrality.

5. Stock Centrality and Optimal Portfolio Weights: In-Sample Evaluation

From section 2, our theoretical results clearly state the importance of the interaction between the individual and the systemic performance of stocks in determining their optimal portfolios weights. This section attempts to empirically disentangle such interaction by means of an in-sample approach. In doing so, we rely on the dataset d-S&P500 which was described in section 3. We further divide this section into two subsections. Subsection 5.1 reports a cross-sectional analysis that uses the complete dataset. In subsection 5.2, we introduce a time series approach addressed by a rolling window analysis.

5.1 Cross Sectional Approach

The complexity characterizing any stocks in capital market is well captured by $\hat{\Omega}$. Minimum Spanning Trees (MST), firstly introduced by (Mantegna 1999), have been intensively used in order to reduce such complexity by efficiently capturing the skeleton of the market through a network structure. In simple terms, MST connects all of the $n$ stocks in the market in a tree network by considering only the highest $n-1$ correlations of returns as links as long as they don’t form any

Figure 5: Average percentage of centrality switching for low, medium and high central stocks
loop in the network\(^6\). This technique has been successfully applied in different country-specific market in order to uncover their hidden backbone, see (Bonanno et al. 2003), (Onnela et al. 2003b), for the US market, (Jung et al. 2006) for the Korean market, (Garas & Argyrakis 2007) for the Greek Market, (Huang et al. 2009) for the Chinese market and (Namaki et al. 2011) for the Iranian market.

Figure 6 depicts the MST for our dataset where the node’s sizes are scaled by the optimal portfolio’s weights in \(m\text{-var}\) strategy given by equation (4). The intensity of the node’s color represents their eigenvector centrality (more intense color corresponds to higher centralities scores). The pattern formalized in corollary 1 where highly central stocks tend to receive lower amount of wealth in an optimal portfolio is clearly seen in figure 6. Therefore, investor’s capital (size of the nodes) tends to be optimally allocated toward lighter nodes placed in the periphery of the network.

![Figure 6: MST for the 200 most capitalized companies of the S&P-500. The size of nodes corresponds to the optimal weights from equation (4) and the intensity of the color regards to its eigenvector centrality.](image)

In order to better understand the interaction between the stocks individual and systemic performances for different investment strategy, let us consider the next two figures. In figure 7, the horizontal axis measures the eigenvector centrality for each of the stocks and the vertical axis the corresponding standard deviation of return. The size and color of the bubbles account for the optimal weights assigned to each stock for \(m\text{-var}\) strategy\(^7\). This figures shows that most of the

---

\(^6\) In a closely related approach, (Onnela et al. 2003a) introduces the Asset Graph (AG) whose algorithm follows the same steps without the restriction of constructing a tree network.

\(^7\) In the calculation of optimal weights for both, min and mean variance strategies, short sales are allowed. Therefore, negative weights are possible
investor’s wealth is allocated in the bottom-left corner of the graph in a set of stocks with relatively low standard deviation of returns (good individual performance) and low eigenvector centrality (good systemic performance).

Similar to figure 7, figure 8 presents the optimal portfolios weights when the $M$-var strategy is considered for a low (left panel) and a high (right panel) portfolio’s returns requirements. Low and high portfolios returns requirement are set equal to 10% and 40% of the maximum possible return reachable in this dataset. For low returns requirements, stocks showing middle levels of Sharpe ratio absorb much of the weights. As it was expected, as long as the returns requirement increases, higher Sharpe ratio stocks receive more weights. However, it should be noted that in any of the cases, the strategy tends to optimally avoid the allocation of wealth towards the center of the stock market network, say stocks with centrality larger than 0.08.
For getting a greater insight about the optimal selection process and its interaction with the individual and systemic performance of stocks, table 4 is provided. This table contains the results of the OLS estimation of the next two equations where the optimal portfolio’s weight of asset \(i\) is a function of its centrality and the standard deviation of returns or the Sharpe ratio depending on the corresponding strategy\(^8\).

\[
\begin{align*}
    w_{i, gmv}^* &= \beta_0 + \beta_1 Centrality_i + \beta_2 Std_i + \epsilon_i \\
    w_i^* &= \beta_0 + \beta_1 Centrality_i + \beta_2 Sharpe Ratio_i + \epsilon_i
\end{align*}
\]  

\[
\begin{array}{c|c|c|c|c|c}
\text{Optimal Weights} & \text{Centrality} & \text{Std} & \text{Sharpe Ratio} & \text{N} & \text{R}^2 \\
\hline
\text{Min Variance} & -0.740 & -1.755 & & 200 & 0.231 \\
\hline
\text{Mean Variance} & -0.778 & & 0.761 & 200 & 0.179 \\
\hline
\end{array}
\]

Table 4 * p<0.05, ** p<0.01, *** p<0.001

As it was expected, for both, minimum variance and mean variance strategy, higher values of centrality implies lower values of weights in an optimal portfolio. This accounts for the systemic effect of the stocks. The first columns of table 5 shows this result in which the estimated coefficient \(\beta_1\) assumes a negative and statistically significant values. The effect of individual stocks as if they were in isolation is captured by the coefficient \(\beta_2\), corresponding to the response of optimal weights to standard deviation of return and to the Sharpe ratio for the \(m\)-var and \(M\)-var strategy, respectively.

In equation (12), \(\beta_2\) assumes negative and significant values pointing out how the optimal strategy discourages the inclusion of highly volatile stocks into the portfolio. In equation (13), \(\beta_2\) assumes a positive and significant values, indicating how the optimal strategy strengthens the inclusion of stocks with individual high Sharpe ratio into the portfolio.

### 5.2 A Time Series Approach

Throughout this section, the entire data set is divided into 2522 overlapping rolling windows of 60 days of length (one quarter approximately). The chosen overlapping segment is 59 days since each new window under analysis is constructed by a one-day displacement step. Let us define the mean stock centrality \((MC)\) and the weighted stock centrality \((WC)\) as follows:

\[^8\] In equation 13, portfolio’s weights are obtained assuming a required return equal to 40% of the maximum possible return reachable with such dataset.
\[
MC_t = \frac{1}{n} \sum_{i=1}^{n} Centrality_{it} 
\]

(14)

\[
WC_t = \sum_{i=1}^{n} W_{it}^* Centrality_{it} \text{ for } W_{it}^* = w_{it,mv} \text{ and } w_{it}^*
\]

(15)

Therefore, when the weighted centrality gets closer to the mean centrality, the optimal strategy assigns wealth roughly equal to \(1/n\) rule. On the other hand, when the weighted centrality is lower (higher) than the mean centrality, the optimal strategy allocates wealth toward the periphery (center) of the network. Figure 9 presents these two measures for both of the considered investment strategies through time.

![Figure 9: Mean Centrality and Optimally Weighted centrality for m-var strategy (left) and M-var strategy (right).](image)

It should be highlighted that the weighted centrality tends to be lower than mean centrality most of the time. Therefore, the best allocation of resource is placed toward the outskirt of the network which is consistent with (Pozzi et al. 2013). However, there are period of times where the weighted centrality gets closer to the mean centrality. This time-dependent behavior could be explained due to the correlation between individual performance and the systemic performance of the assets in the stock market network.

Take the case of \(m\text{-}var\) strategy. Suppose that the set of stocks with the lowest standard deviations also corresponds to the set of stocks with the lowest centrality. In this case, investing in non-central assets is for sure the best alternative to follow. On the other hand, when the lowest central assets are also the ones with the highest volatilities, a trade-off appears and an optimal portfolio balances these two dimensions. In other words, when the correlation between the standard
deviation of return and centrality is positive, investing toward the periphery is optimal. We call this correlation as $\pi$. However, when $\pi$ is negative, it is no longer true that peripheral stocks are the unconditional best destination of wealth. For the case of $M$-var strategy, we define the correlation between Sharpe ratio and centrality of stocks as $\rho$. Following the same logic as before, for $\rho < 0$, the highest Sharpe ratios are placed in the periphery of the network. Therefore no trade-off exists and wealth should be allocated in the outskirt of the market. As long as $\rho$ assumes positive values, highly central stocks are also the ones with the best individual performances and, again, an optimal balance is required. Figure 10 shows the behavior $\rho$ and $\pi$ through time.

![Figure 10: Correlation between the individual performance and the systemic performance. (Left)- Between Sharpe ration and Centrality ($\rho$). (Right)- Between Standard deviation of return and Centrality. ($\pi$)](image)

The essential message from this section is that the target region of the market network from which to select assets depends crucially on the interaction between the systemic and individual dimension, says it crucially depends on $\rho$ or $\pi$ (conditional on the relevant strategy under consideration). Further, such correlations are not static but they change through time (and most surely across markets). With the aim to clarify this point, table 5 presents OLS estimation of the equation (16) and (17) where $WC_{t,mv}$ and $WC_{t,mean}$ stands for the weighted centrality under the $m$-var and $M$-var weights in period $t$, respectively. We include as explanatory variables two measures characterizing the distribution of centrality in period $t$, $MC_t$ accounting for the mean centrality and $CV_{Cent_t}$ standing for its coefficient of variation. Further, in order to take into account the interaction between the individual and systemic stock’s dimensions, we include $\pi_t$ and $\rho_t$ as additional independent variables.
\[ WC_{t, m_{MV}} = \beta_0 + \beta_1 MC_t + \beta_2 CV_{Cent} + \beta_3 \pi_t + \varepsilon_t \quad (16) \]
\[ WC_{t, mean} = \beta_0 + \beta_1 MC_t + \beta_2 CV_{Cent} + \beta_3 \rho_t + \varepsilon_t \quad (17) \]

Table 5 shows that an increase in mean centrality pushes the investment allocation towards the periphery of the network for both strategies. Therefore, when the mean centrality in the network rises, the potential advantage of moving away from the center increases its reward. Considering the dispersion of the centrality distribution, we conclude that when the coefficient of variation of such distribution increases (controlling for the mean centrality), there is more room for diversification. This allows optimal investors to avoid the negative effects of stock with poorly systemic performance by moving farther away from the center of the network. With regards to \( \pi \) and \( \rho \), the signs of the coefficients \( \beta_3 \) in table 5 are the ones we should expect. For the \( m\text{-var} \) case, it is negative which implies that when \( \pi > 0 \), the least central nodes also show the lowest standard deviation of return, thus investors should reduce the allocation of resources in the center toward the periphery of the network. For the \( M\text{-var} \) case, the sign of \( \beta_3 \) is positive. When \( \rho > 0 \), high central stocks are associated with high Sharpe ratios and then investor wealth should take this effect into account by moving away from the periphery toward the center of the network. Therefore, \( \beta_3 \) in equation (16) and (17) relates to the optimal location of the target region to invest. Such a region is conditionally placed in accordance to states of the market which are properly captured by \( \pi \) and \( \rho \).


According to (DeMiguel et al. 2009), naïve strategy, commonly named as \( 1/N^9 \), is shown not to be consistently outperformed by neither \( M\text{-var} \) strategy nor its extensions designed to deal with estimation error problem. This better out-of-sample performance of \( 1/N \) strategy relative to Markowitz’s rule is also investigated and supported by (Jobson & Korkie 1980), (Michaud 2008) and (Duchin & Levy 2009). In this section, we examine the extent to which network-related strategies can be used in order to improve the out-of-sample portfolio by investing in a target group of assets.

---

\( ^9 \) Naïve strategy assigns a fraction \( 1/n \) of wealth to each asset out of the \( n \) available assets.
Based on section 5, this target group of assets results from the removal of stock from opportunity set in accordance to their centrality.

6.1 The Out-of-Sample Evaluation Approach

The out-of-sample evaluation is as follows. A “rolling window” scheme is adopted for generating out-of-sample returns. Let us suppose that we have in total $T$ periods of stock’s returns in the dataset. In the first step, an $M$-period estimation window is set up. This sample of $M$ periods is used to calculate eigenvector centrality of stocks from the shrinkage covariance matrix (Ledoit & Wolf 2004). With this measure of centrality at hand, we discard stocks in accordance to specific stocks removal rules, which are going be defined below, leading to a target group of stocks. Then, the portfolio to be hold for the next period, $w_{M+1}$, results from applying naïve diversification upon the mentioned set of target stocks. This process is repeated by adding the return for the next $H$ data points and dropping the $H$ earliest ones to the $M$-period sample, until the end of the dataset is reached. At the end, we end up with $(T - M)/H$ vectors of portfolio’s weights for each of the stock’s removal rule.

Given the empirical results commented at the beginning of this section, we consider as the benchmark strategy to beat, the naïve rule applied upon all of the available stocks in the opportunity set. We propose $\rho$-dependent strategy as an alternative network-based investment rule working as follows. In each rebalancing period, $\rho$ is calculated and then 20 stocks are selected in accordance to its value: If $\rho$ is sufficiently high, $\rho > \bar{\rho}$, we invest in the 20 highest central stocks; if $\rho < \bar{\rho}$, we invest in the 20 lowest central stocks where $\bar{\rho}$ accounts for a threshold to be specified. Note that in applying $\rho$-dependent strategy, we do not have full information about future realizations of $\rho$. Instead, we make a decision about the target region of the market network to invest in accordance to its current value. To make sure that the performance of $\rho$-dependent strategy is not made by chance, we also take into account the reverse $\rho$–dependent strategy working in reverse order: If $\rho > \bar{\rho}$, we invest in the 20 lowest central stocks and if $\rho < \bar{\rho}$, we invest in the 20 highest central ones. As a sort of a “control strategies”, we consider three additional unconditional rules that naively invest in the 20 stocks showing the Highest Sharpe Ratio, the Highest Centrality and the Lowest Centrality. It is important to mention that the decision of constructing portfolios made up by 20 assets is based on (Desmoulins-Lebeault & Kharoubi-Rakotomalala 2012) which highlights the fact that most of the diversification benefits is gained by investing in such number of stocks.

The strategies are compared using three out-of-sample portfolio performance measures: $i)$ Sharpe ratio, $ii)$ variance of return and $iii)$ turn-over. This last measure captures the amount and sizes of the operations that are required for rebalancing purpose and it is calculated as follows:

$$\text{turnover} = \frac{1}{T - M - 1} \sum_{t=M}^{T-1} \sum_{j=1}^{N} |w_{j,t+1} - w_{j,t}|$$  \hspace{1cm} (18)
where \( w_{jt+1} \) is the weight for asset \( j \) at the beginning of year \( t+1 \) and \( w_{jt+} \) is the weight of asset \( j \) before rebalancing between the end of year \( t \) and beginning of year \( (t+1) \). For an statistically assessment of the difference in performance between the benchmark strategy and the rest of portfolio strategies, we follow (Ledoit & Wolf 2008) and (Ledoit & Wolf 2011) when Sharpe ratio and portfolio variance is under analysis. The calculation of p-values are based on studentized circular block bootstrap with block size equal to 5 and number of bootstrap samples equal to 5,000. For the particular case of the variance, stationary bootstrap is employed as in (Politis & Romano 1994).

**6.2 Defining the value of \( \bar{\rho} \)**

Before moving on to present the results of the out-of-sample analysis, the value of \( \bar{\rho} \) must specified. For negative values of \( \rho \) no trade-off exists. The stocks from the periphery show the largest Sharpe ratio. As a consequence, the investor’s wealth should be allocated toward the outskirt of the network. When \( \rho \) assumes values close to zero, there is no tendency for the stocks with the highest individual performance to be located in any particular region of the market network. Therefore and in order to avoid large portfolios variances, the periphery must also be the prefer region to invest. As long as \( \rho \) assumes positive values, the trade-off between the individual and the systemic dimension of stocks emerges; then moving the target region toward the center of the market network comprises a beneficial choice.

To understand the role of \( \bar{\rho} \), we artificially create different data sets showing \( \rho \) that span over broad range of values. We rely on a subsample of \( d \)-NYSE dataset for this purpose (see appendix D). In particular, we randomly generate 120 data sets of 150 stocks each with \( \rho \) values ranging from -0.20 to 0.45. With the aim to calculate the out-of-sample Sharpe ratio, M and H were set be equal to 500 days and 20 days, respectively. Figures 11 and 12 present the results of the portfolio’s Sharpe ratio for several sub-baskets of stocks constructed by a progressively deleting some of them from the tails of the centrality distribution (with 5% increments). Figure 11 accounts for \( \rho \) approximately equal to zero while figure 12 considers the extreme cases of \( \rho \) equal to 0.45 (left panel) and -0.20 (right panel).

![Figure 11: Out-of-sample Sharpe ratio (colors) for different stocks removal of artificially created datasets with \( \rho =0.000 \)](image)
Figure 11 present a scenario in which intensively removing high central stocks leads to better out-of-sample performance in terms of Sharpe ratio. This result is expected since highly individual performing stocks are randomly disseminated in the market network. Therefore, the preferable choice points to the periphery to obtain the best results. The right panel of figure 12 (for $\rho = -0.20$) also shows us the convenience of selecting the periphery as the target region to invest. This is due to the absence of trade-off between the systemic and individual dimension of stocks. The opposite case is observed in the left panel of figure 12 (for $\rho = 0.45$). In this case, the best portfolio’s performances are achieved by investing in high central stocks with an intense asset’s removal from the left tail of the centrality distribution. In this case the positive effect of the individual dimension of stocks over-compensates the negative effect of their systemic performance.

The critical question arising regards to the break point of $\rho$ that defines the region of the market network that is susceptible to be discarded from the investment opportunity set. In this regard, we consider the following sensitivity test. Let us define the high central stock investing region as the set of stocks arising from the deletion of 25% to 45% of assets from the left tail of the centrality distribution and no more than 20% from its right tail. In a symmetric fashion, let us define the low central stock investing region as the set of stocks comprising the deletion of 25% to 45% of stock from the right tail of the centrality distribution and no more than 20% of its left tail. Then, from all of the 120 artificially-constructed data sets, we identify the investing region that generates the highest out-of-sample Sharpe ratios and the corresponding $\rho$. The identification rule simply averages out the out-of-sample Sharpe ratios generated in each one of the two commented regions and then selecting the one with the highest average performance. The results are depicted in figure 13 showing the distribution of $\rho$ conditional to the region generating the largest Sharpe ratio.
From the above graph, it could be inferred that low values of $\rho$ are more consistent with outperforming Sharpe ratio coming out from the *low central stock investing region*. Contrary to this, for large values of $\rho$, it is the *high central stock investing region* the one that tends to generate the highest risk-adjusted returns. To conclude this section, we need to set up a positive value of $\rho$ that must be high enough in order to be worthwhile to move the optimal investment region from the outskirt of the market network towards investing in highly central stocks. In this regard and taking figure 13 into account, we consider as reasonable value for $\tilde{\rho}$ to be equal to 0.2 since this values roughly coincides with the 75% percentile of $\rho$ from the *low central stock investing region* and the 25% percentile of $\rho$ from the *high central stock investing region*. Determining $\tilde{\rho}$ is the weakest point in our procedure since it implies ad-hoc rules. What it must be clear, $\tilde{\rho}$ should be large enough in order to be worth to move the investment region toward the center of the market network.

6.3 Out-of-Sample Performance of $\rho$–dependent strategy

In what follows, we rely on three different datasets to run our experiments. The first one corresponds to *d-S&P500* (accounting for the 200 most capitalized stocks in S&P500) described in section 3. The second one regards to daily stock returns of the 200 most capitalized stocks in FTSE 250 index from Feb 2006 until Oct 2013. We called this dataset as *d-FTSE250*. The third dataset could be considered as a subset of *m-NYSE* dataset since the monthly returns of the 202 firms that remains alive from Jan 1964 to Dec 2006 were taken into account. We named this dataset as *modified m-NYSE* (see appendix D for a detailed description of data sets). For the out-of-sample performance analysis, we consider $M=1.000$ and $H=20$ for the daily datasets and $M=192$ and $H=12$ for the monthly dataset. As commented previously, $\tilde{\rho} = 0.2$.

The results of the experiment are presented in table 6. It is noticeable the outperformance of $\rho$–dependent strategy with daily datasets. The Sharpe ratio from investing in all stocks using naïve strategy reaches to 0.471 and to 1.153 for S&P 500 and FTSE 250 datasets, respectively. The same measure rises to 0.724 and 2.584 when $\rho$–dependent strategy is in place showing a statistically
significance difference with respect to the benchmark. Additionally, we can see that in both datasets, investing in accordance to \( \rho \)-dependent strategy lead to a significant lower portfolio’s variance. It reduces from 0.066 to 0.051 and from 0.025 to 0.015, for the cases of S&P 500 and FTSE 250, respectively. Considering the NYSE monthly dataset, the \( \rho \)-dependent results does worst in terms of Sharpe ratio and variances in comparison to the benchmarks, however, the p-values indicate a non-significant difference. Additionally, table 6 also shows that our \( \rho \)-dependent strategy implies a distinctive and enhanced investment dynamics when it is compared to the unconditional Highest Sharpe ratio or Lowest Central strategies. In general terms these two control strategies present poorer outcomes in portfolio’s Sharpe Ratio for all of the considered datasets. In order to discard the possibility that our results were driven by chance, table 6 also reports the results for the reverse \( \rho \)-dependent. In this case, none of the portfolio’s performance measures show better results when compared either with the benchmark or with \( \rho \)-dependent strategy. Unfortunately, the major shortcoming of our proposed investment strategy is that it raises the amount of rebalancing operations, a phenomenon captured by the increased portfolio’s turnover, compared to the benchmark strategy.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P 500 (Avrg ( \rho ): -0.0556)</th>
<th>FTSE 250 (Avrg ( \rho ): -0.1853)</th>
<th>NYSE (Avrg ( \rho ): 0.1157)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sharpe Ratio</td>
<td>Variance</td>
<td>Turnover</td>
</tr>
<tr>
<td>All Stocks</td>
<td>0.471</td>
<td>0.066</td>
<td>0.141</td>
</tr>
<tr>
<td>( \rho )-dependent</td>
<td>0.724 (0.0125)</td>
<td>0.051 (0.0010)</td>
<td>0.149 (0.008)</td>
</tr>
<tr>
<td>Reverse ( \rho )-dependent</td>
<td>0.315 (0.1520)</td>
<td>0.110 (0.0009)</td>
<td>0.126 (0.009)</td>
</tr>
<tr>
<td>Highest Sharpe Ratio</td>
<td>0.438 (0.8239)</td>
<td>0.038 (0.0009)</td>
<td>0.141 (0.009)</td>
</tr>
<tr>
<td>Highest Central</td>
<td>0.437 (0.1920)</td>
<td>0.038 (0.0009)</td>
<td>0.126 (0.009)</td>
</tr>
<tr>
<td>Lowest Central</td>
<td>0.470 (0.2265)</td>
<td>0.077 (0.0009)</td>
<td>0.149 (0.009)</td>
</tr>
</tbody>
</table>

Table 6: Out-of-sample performance of portfolio strategies (annualized values) and \( \hat{\beta} = 0.3 \).

Two additional comments are worth to be mentioned. First, note that the relatively extreme and negative value of \( \rho \) (-0.1853 on average) for the UK market could help to explain the striking good results obtained with such database. When the monthly dataset of active NYSE’s firms is considered, the significance of our results is undermined. One possible reason for this to happen relates to the existing trade-off between individual performance and systemic performance given its positive \( \rho \) (0.1157 on average). Second, it should be mentioned that the prescription from (Pozzi et al. 2013) of unconditionally investing only in the periphery of the network shows results in-line or worsen than the benchmark but clearly inferior to the ones from \( \rho \)-dependent strategy.
7. Conclusion and Future Research lines

In this paper, the capital market is interpreted as a network where the stocks correspond to nodes and the links between two of them relates to their pair correlation of returns. We prove theoretically the tendency to include stocks located in periphery of such a structure when mean-variance or minimum-variance strategies are in place. In other words, Markowitz’s rule discourages the inclusion of central firms in an optimally designed portfolio.

From a more descriptive point of view, we find that both, financial and market variables, are major determinants of stock’s centrality. Additionally, we investigate the relative stability of listed firms with respect to their place in the network. The results highlight the higher likelihood for low central stocks to be delisted from the market when it is compared to the rest of the assets. Further, considering only the assets that remain listed, we present evidence supporting the tendency for them to maintain their position in the network through long periods of times.

Next, we investigate, by means of in-sample and out-of-sample approach, the performance of network-based investment strategies. We introduce $\rho$ as the correlation between the individual performances of stocks, Sharpe ratios, and their systemic performance which is properly captured by their centrality. Our out-of-sample results show that higher portfolio’s Sharpe ratios and lower portfolio’s variance could be achieved by a proper specification of the so-called $\rho$–dependent strategy. This strategy results from the elimination of stocks from the investment opportunity set in accordance to the value of $\rho$. For low values of $\rho$, naïve strategy is applied only to a set of stocks belonging to the periphery of the market network while for $\rho$ larger than a threshold $\tilde{\rho}$, the target region moves toward a set of more central nodes. As it was commented, the major weaknesses of the approach regards to the definition of $\tilde{\rho}$.

We strongly believe in using network theory as a suitable device to improve the portfolio selection process. The current paper is an attempt to contribute in such a direction by focusing in a particular portion of the market to be the preferred destination for investor’s wealth. As a future research line, it seems interesting to investigate whether this approach could be embedded into a directed network framework. This suggestion presents the benefits to come out with an innovative approach to be applied for stress-testing in the stock markets and to clearly identify the roles of stocks as shock-absorber or shock booster.

REFERENCES


APPENDIX A – Proof of Proposition 1

Preliminary Results

Let us assume two square $n \times n$ matrices $\Omega_0$ and $\Omega_1 = \Omega_0 + a \cdot I$ where $a$ is a scalar and $I$ is the identity matrix. The sets of eigenvectors for these matrices are given by $\{v_1^0, ..., v_n^0\}$, $\{v_1^1, ..., v_n^1\}$ and the related sets of eigenvalues by $\{\lambda_1^0, \lambda_0^0\}$ and $\{\lambda_1^1, \lambda_1^1\}$ where the supra-indices 0 or 1 refers to the corresponding matrices. As a preliminary result, we show that the eigenvectors of $\Omega_0$ and $\Omega_1$ are exactly the same, say $\{v_1, ..., v_n\}$ and the corresponding associated eigenvalues are related as follows: $\lambda_k^1 = \lambda_k^0 + a$ for $k = 1 \ldots n$.

By definition, we know that $\lambda_k^1 v_k = \Omega_1 v_k$ and $\lambda_k^0 v_k = \Omega_0 v_k$ then

\[
\lambda_k^1 v_k = \Omega_1 v_k
\]
\[
\lambda_k^1 v_k = (\Omega_0 + a \cdot I) v_k
\]
\[
\lambda_k^1 v_k = \Omega_0 v_k + a \cdot v_k
\]
\[
\lambda_k^1 v_k = \lambda_k^0 v_k + a \cdot v_k
\]
\[
\lambda_k^1 v_k = (\lambda_k^0 + a) v_k
\]

Therefore

$\lambda_k^1 = \lambda_k^0 + a$

Proof of Proposition 1

Let us move on to prove Proposition 1 for the mean-variance case. The result from the minimum variance strategy follows exactly the same logic and therefore its derivation is not exposed except for its final result. The correlation matrix denoted by $\Omega$ is a $n \times n$ diagonalizable symmetric matrix with $\{v_1, ..., v_n\}$ as its set of eigenvectors and $\{\lambda_1, ..., \lambda_n\}$ as the set of the corresponding eigenvalues arranged in descendent order. Then, we can write $\Omega = P \Lambda P^T$ where $P$ is an $n \times n$ orthogonal matrix containing the eigenvectors of $\Omega$ in columns $v_1, ..., v_n$ and $\Lambda = diag(\lambda_i)$. It follows that:

\[
\Omega^{-1} = P \Lambda^{-1} P^T = P \cdot diag(1/\lambda_i) \cdot P^T
\]

A.1

\[
\Omega^{-1} = [v_1, ..., v_n] \begin{bmatrix}
1/\lambda_1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1/\lambda_n
\end{bmatrix} [v_1^T, ..., v_n^T]^T
\]

A.2
\[ \Omega^{-1} = \sum_k \left( \frac{1}{\lambda_k} v_k v_k^T \right) = \frac{1}{\lambda_1} v_1 v_1^T + \frac{1}{\lambda_2} v_2 v_2^T + \cdots + \frac{1}{\lambda_n} v_n v_n^T \]  

A.3

From equation (10) in section 2.2, we have

\[ \hat{\omega}^* = \varphi \Omega^{-1} \hat{\mu}^e \]  

A.4

\[ \hat{\omega}^* = \varphi \Omega^{-1} \hat{\mu}^e + \varphi \hat{\mu}^e - \varphi \hat{\mu}^e \]  

A.5

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi [\Omega^{-1} \hat{\mu}^e - \hat{\mu}^e] \]  

A.6

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi [\Omega^{-1} - I] \hat{\mu}^e \]  

A.7

Using A.3 and the preliminary results from this appendix, we know that the matrix \( \Omega^{-1} - I \) presents the same eigenvectors with eigenvalues equal to \( \frac{1}{\lambda_k} - 1 \) for \( k = 1 \ldots n \). Then

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi \left[ \sum_k \left( \frac{1}{\lambda_k} - 1 \right) v_k v_k^T \right] \hat{\mu}^e \]  

A.8

Give that eigenvector centrality regards only to the eigenvector corresponding to the largest eigenvalue we define \( \Gamma = \varphi \left[ \sum_{k=2}^{n} \left( \frac{1}{\lambda_k} - 1 \right) \right] \hat{\mu}^e \). Then, A.8 could be rewritten as

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi \left( \frac{1}{\lambda_1} - 1 \right) v_1 v_1^T \hat{\mu}^e + \Gamma \]  

A.9

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi \left( \frac{1}{\lambda_1} - 1 \right) (v_1^T \hat{\mu}^e) v_1 + \Gamma \]  

A.10

Since \( v_1^T \hat{\mu}^e \) accounts for the expected return of the market from the Principal Component perspective, we define \( \hat{\mu}_M = v_1^T \hat{\mu}^e \)

\[ \hat{\omega}^* = \varphi \hat{\mu}^e + \varphi \left( \frac{1}{\lambda_1} - 1 \right) \hat{\mu}_M v_1 + \Gamma \]  

A.11

For the case of minimum variance strategy, the final equation is as follows

\[ \hat{\omega}^* = \varphi_{mv} \hat{\mu}^e + \varphi_{mv} \left( \frac{1}{\lambda_1} - 1 \right) \epsilon_M v_1 + \Gamma_{mv} \]  

A.12
Where $\epsilon_M = (v_1^T \epsilon)$ standing for the inverse standard deviation of returns at market level and 

$I_{mv} = \varphi_{mv} \left[ \sum_{k=2}^{n} \left( \frac{1}{\lambda_k} - 1 \right) \right] \epsilon$

**APPENDIX B – Descriptive Statistics of Stocks Performance in terms of centrality**

<table>
<thead>
<tr>
<th>Sharpe Ratio</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Q5%</th>
<th>Q95%</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>3.31%</td>
<td>3.03%</td>
<td>1.57%</td>
<td>-0.22%</td>
<td>7.92%</td>
<td>1.23%</td>
<td>6.16%</td>
<td>0.51</td>
<td>0.13</td>
</tr>
<tr>
<td>Medium</td>
<td>2.82%</td>
<td>2.85%</td>
<td>1.13%</td>
<td>-0.07%</td>
<td>5.86%</td>
<td>0.88%</td>
<td>4.35%</td>
<td>-0.02</td>
<td>-0.06</td>
</tr>
<tr>
<td>High</td>
<td>2.88%</td>
<td>2.86%</td>
<td>0.92%</td>
<td>1.03%</td>
<td>5.33%</td>
<td>1.36%</td>
<td>4.04%</td>
<td>0.00</td>
<td>-0.66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Excess Return</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.07%</td>
<td>0.06%</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>0.05%</td>
<td>0.03%</td>
<td>0.02%</td>
</tr>
<tr>
<td></td>
<td>-0.01%</td>
<td>0.00%</td>
<td>0.02%</td>
</tr>
<tr>
<td></td>
<td>0.22%</td>
<td>0.13%</td>
<td>0.16%</td>
</tr>
<tr>
<td></td>
<td>0.02%</td>
<td>0.02%</td>
<td>0.03%</td>
</tr>
<tr>
<td></td>
<td>0.31%</td>
<td>0.12%</td>
<td>0.09%</td>
</tr>
<tr>
<td></td>
<td>1.20%</td>
<td>0.58%</td>
<td>0.85%</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>-0.08</td>
<td>2.59</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviation</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.05%</td>
<td>2.03%</td>
<td>2.24%</td>
</tr>
<tr>
<td></td>
<td>0.70%</td>
<td>0.63%</td>
<td>0.50%</td>
</tr>
<tr>
<td></td>
<td>1.09%</td>
<td>1.09%</td>
<td>1.43%</td>
</tr>
<tr>
<td></td>
<td>4.84%</td>
<td>3.82%</td>
<td>3.76%</td>
</tr>
<tr>
<td></td>
<td>1.20%</td>
<td>1.15%</td>
<td>1.49%</td>
</tr>
<tr>
<td></td>
<td>3.21%</td>
<td>3.24%</td>
<td>3.17%</td>
</tr>
<tr>
<td></td>
<td>1.30</td>
<td>0.58</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>2.53</td>
<td>-0.23</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table B.1

**APPENDIX C – Correlation and Centrality for the 200 most capitalized companies of the SP500- Daily returns**

![Figure C1: Distribution of correlation](image)

<table>
<thead>
<tr>
<th>Descriptive Statistics - Return Correlation</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
<th>Min</th>
<th>Max</th>
<th>Q5%</th>
<th>Q95%</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Descriptive Statistics - Return Correlation</td>
<td>0.4131</td>
<td>0.4004</td>
<td>0.0947</td>
<td>0.1185</td>
<td>0.8694</td>
<td>0.2815</td>
<td>0.5859</td>
<td>0.7225</td>
<td>0.9982</td>
</tr>
</tbody>
</table>

Table C.1
### APPENDIX D – Data Bases

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source</th>
<th>Num of Firms</th>
<th>Frequency</th>
<th>Periods</th>
<th>Filters and Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF RF</td>
<td>Risk free rate</td>
<td>Kenneth R. French web page</td>
<td>-</td>
<td>Daily</td>
<td>01/07/1926 – 31/12/2013 23133</td>
<td>To be used in calculating excess returns</td>
</tr>
<tr>
<td>d-S&amp;P500</td>
<td>daily excess returns (split and divided adjusted) for the 200 most capitalized stocks integrating the SP500 index in 2013</td>
<td>CRSP 200</td>
<td>Daily</td>
<td>01/10/2002 – 31/12/2012 2581</td>
<td>1- Firms with return's information for every period 2- 200 most capitalized companies in 2012 3- No negative equity in any of the periods 4- SIC code also taken from this data set</td>
<td></td>
</tr>
<tr>
<td>d-NYSE</td>
<td>Stock prices of firms listed in NYSE after normalization for capitalization and cheap stocks</td>
<td>CRSP 947</td>
<td>Daily</td>
<td>06/01/2004 – 30/07/2007 2010</td>
<td>The stocks are chosen to have capitalization more than 20th percentile of market capitalization and prices higher than $5.</td>
<td></td>
</tr>
<tr>
<td>S&amp;P 500 COMPSTAT</td>
<td>financial variable of the companies in SP500 CRSP</td>
<td>COMPSTAT 200</td>
<td>Quarterly</td>
<td>QI-2003 – QIV-2012 40</td>
<td>The next items were included: 1- NIQ: Net income 2- ATQ: Total Assets 3- CHEQ: Cash and Short Term Assets 4- LTQ: Total Liabilities 5- MKVAQ: Market value of the company 6- CSTTRQ: Volumes traded of stock 7- SEQQ: Total Equity</td>
<td></td>
</tr>
<tr>
<td>d-FTSE250</td>
<td>Adjusted Prices and Return of components of FTSE 250</td>
<td>Datastream 200</td>
<td>Daily</td>
<td>27/02/2006 – 15/10/2013 1991</td>
<td>1- Rf rate taken as UK TREASURY BILL TENDER 3M - MIDDLE RATE</td>
<td></td>
</tr>
<tr>
<td>m-NYSE</td>
<td>Stocks from NYSE after normalization for capitalization and cheap stocks</td>
<td>CRSP All listed companies</td>
<td>Monthly</td>
<td>Apr-68 – Apr-12</td>
<td>Only those firms that were alive thought all the periods</td>
<td></td>
</tr>
<tr>
<td>modified m-NYSE</td>
<td>Stocks from NYSE after normalization for capitalization and cheap stocks (Daily returns)</td>
<td>CRSP 202</td>
<td>Monthly</td>
<td>Jan-64 – Dec-06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure C2: (Left) Distribution of centrality. (Right) Scatter plot between centrality and optimal weights for the minimum variance strategy.