A Network Approach to Unravel Asset Prices Co-movements Using Minimal Dependence Structures

Pablo Jose Campos de Carvalho

Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY

Banco Central do Brasil, Brasilia, Brazil *

Aparna Gupta

Lally School of Management, Rensselaer Polytechnic Institute, Troy, NY

January 13, 2016

*The views expressed in this work are those of the authors and do not reflect those of the Banco Central do Brasil or its members.

A Network Approach to Unravel Asset Prices Co-movements Using Minimal Dependence Structures

Abstract: In many fields, networks have been used to filter information and describe connected systems. In this paper, we build on the minimum spanning tree (MST) literature developing a layered MST that uses a multi-factor model to explain the dynamic dependencies among elements using systematic and idiosyncratic components of asset prices. This framework proves to be flexible with changes in the underlying data and the choice of factors for the investigation. We show applications of our framework in different contexts and observe that the methodology is helpful in understanding the change of the interdependencies among entities in a data-set. Using this approach we are able to demonstrate dramatic changes in the topology of asset prices networks.

1 Introduction

The interdependence between elements of a financial set is central for decision making and risk management. From the classical portfolio theory investigating the benefits of diversification to present issues in finance such as monitoring and managing systemic risk, the estimation and interpretation of the co-dependencies is very important for financial management. However, identifying the co-dependencies poses many difficulties, especially when the number of variables is large or co-dependence is time-varying, once estimating parameters on the order of the square of the number of assets leads to too many spurious inferences. Focusing on a global-level description, viewing the financial set as complex system, can be a complementary tool to risk management.

In many practically important areas, a network representation significantly aids the analysis of large datasets (Huang et al., 2009). A network representation of co-dependencies requires defining a topological space of nodes and links defining interconnections and interactions between nodes. In this paper, we develop a framework for the interpretation of temporal aspects of co-dependencies in asset price movements. This framework includes a visual representation of the correlation matrix of returns, that is accomplished using a similarity based network. This network is constructed taking into consideration the relative influence of systematic and idiosyncratic factors and how they change in time. Nodes of this network are connected to others ranked by the strength of their co-movement. We apply this conceptual framework in two different sets of stock prices and provide insights that would not be captured using traditional financial tools.

In modern financial literature, networks can be broadly divided into relation defined networks or similarity-based network. On relation-defined networks the nature of the relationship between the elements is given by some form of event or common capability such as balance sheet claims (Allen and Gale, 2000; Gai et al., 2011) or fund flows (Barro and Basso, 2010; Minoiu and Reyes, 2013). These studies are usually concerned with how stochastic shocks propagate through the network. Similarity-based networks, on the other hand, focus on the similarity between nodes such as correlation of returns (Mantegna, 1999), common contracts (De Masi and Gallegati, 2012), financial statement variables (Papadimitriou et al., 2013) and probability of default (Tabak et al., 2011).

The study of the relationship between diversity and system stability has been the focus of many fields in science (Engle et al., 2010). Similarly, financial markets are also highly interconnected with rather complex relationships between financial instruments. The proposed network rellies on a distance measure based on co-movements of daily returns as links among the elements. Such approach has been extensively used in many different contexts. For instance, Aste et al. (2005) study interest rate data by filtering information from the correlation coefficient matrix to extract a network of the most linked (correlated) elements. A similar approach is taken in Tse et al. (2010), where network properties are analyzed for stock returns from the US equity market. They report that there are a small number of highly connected stocks, and propose that portfolios of these stocks are capable of representing majority of stock market.

Complex systems are often described in hierarchical form, where stronger relationships are first filtered, and specific and minor relations surface later in the process (Simon, 1962). This framework relies on a minimal spanning tree (MST) representation of the network for a minimalist and most significant identification of connections between nodes and their temporal evolution. An MST is a common method to unveil the structure of complex systems (Vandewalle et al., 2001) by progressively built by minimally linking all the nodes together in a graph characterized by the strongest links or smallest distances between the nodes.

Hierarchical clustering, such as MSTs, are filtering procedures useful in multivariate characterization of stock return time series (Tumminello et al., 2010). MSTs were first used in finance by Mantegna (1999) for network analysis of stock price correlations. Using equity prices from the Dow Jones Industrial Average (DJIA) index and the Standard and Poor's 500 (S&P 500) index, Mantegna (1999) constructed a hierarchical arrangement of the stocks after organizing them by their industries. In a further study, Onnela et al. (2003) showed that companies in a minimum risk portfolio are consistently in the farthest areas of the MST, suggesting that even if the hierarchical structure changes in time, these economic clusters can help in portfolio allocation and risk management.

MST framework have also been applied in the context of foreign exchange rates (Bonanno et al., 2004; Mizuno et al., 2006; Naylor et al., 2007) revealing clusters representing geographical regions of the world. The structure, however, is is highly dependent on the numeraire chosen Kwapień et al. (2009), suggesting that there are relevant factors that interferes in the way networks are formed. This is consistent with the evidence that MSTs from one-factor model filter important information about pair-wise dependencies and differ substantially from those constructed directly from the original data (Bonanno et al., 2003).

Our financial network representation incorporates explanatory factors for evaluating the underlying dependency structures of asset nodes. From the development of the Capital Asset Pricing Model (Sharpe, 1964), factor models have been widely used to explain and predict returns. In Fama and French (1993), the authors show that apparent anomalies of the CAPM can be explained with the inclusion of new factors giving rise to multi-factor literature. The use of several factors helps reduce the complexity and randomness of the analysis of the dependence structure between assets. The factors are chosen according to the assets and attempt to reflect common components across assets in their price structure. Thus, these models decompose asset returns in two parts: a systematic component, explained by the chosen factors, and other idiosyncratic, specific to the asset and the orthogonal to the factors.

The topological space in this framework contains three distinct subsets of elements: systematic factors, core asset nodes and idiosyncratic nodes. A core asset node connects with the rest of the network either through one of the explanatory factors or through its idiosyncratic element. In doing this filtration, we focus on understanding the relevance of the underlying explanatory factors on the interdependence between two different assets. Connections arising from two idiosyncratic nodes suggest that similarities between two assets are weakly explained by the factors. This approach is different from the partial correlation in Kenett et al. (2010), where the authors are concerned with the degree to which hedging decision can be made in sectors based on factor regressions. We find that relative and absolute importance of factors change over time.

The choice of the dependence structure is an important feature of studying co-dependence, as the failure to capture interdependence structure or its misidentification can lead to suboptimal and costly decisions (Kole et al., 2007; Chollete et al., 2009). Our approach allows for the use of any interdependence measure in order to define the distance matrix. Without loss of generality, in this paper, we construct distance networks based on Pearson, Spearman, Kendall and exponentially weighted correlation measures.

Beyond the choice of factors, the other challenge in analyzing the dependence structure for asset prices is the temporal modifications in the dependence structure. The misidentification of these co-dependencies can often arise from the fact that these co-dependencies are time-varying (Engle and Figlewski, 2015). The hypothesis of non-stationarity of financial time series has been widely tested, with evidence heavily in favor of non-stationarity of financial time series data. Using data for emerging markets, Boyer et al. (2006) find that co-movement in joint distributions increases during high volatility periods. Non-stationarity may arise from asymmetric volatility spillovers, as seen for Eurozone stock market data (Kohonen, 2013). Similarly, Ang and Chen (2002) study correlations between U.S. stocks and the aggregate U.S. market, where they see clustering behavior that is more pronounced for downside moves, especially for extreme downside moves, than for upside moves. Longin and Solnik (2001) run a similar experiment for large international markets and find no evidence of clustering during periods of high volatility in prices, but still find that correlations are greater during bear markets than in bull markets. Although Forbes and Rigobon (2002) find no significant increase in unconditional correlation coefficients, during the 1997 Asian crisis, 1994 Mexican devaluation, and 1987 U.S. market crash, there were high levels of market co-movements in all these periods, which they term interdependence.

Non-stationary dynamics of asset prices implies that an investor would need to do more than just re-balance portfolios in time, even if their risk-aversion remains unchanged. This would also have implications for the role of explanatory systematic factors changing with time in a multi-factor model. Time-varying beta models are constructed to address this issue (Ang and Chen, 2002). Being able to extract the most essential features for these temporal changes in joint distributions can be highly instructive in how dramatically the portfolio weights must change (Das and Uppal, 2004). In this paper, the financial network is reconstructed over time enabling us to evaluate the temporally evolving co-dependencies between nodes. In a temporally evolving network, although the definition of nodes remains the same, the links between them can change, allowing the examination of relative significance of idiosyncratic versus systematic connections between nodes.

The framework developed in this paper allows analyzing hundreds of asset price variables to minimally identify their interconnections, with careful distinction made to detect the source of connection being idiosyncratic, or through systematic factors. It offers flexibility in the choice of explanatory factors to suit the specific objectives of the study and the context of constructing a minimalist interdependency between nodes. Temporal behavior of the network enhances understanding of the links between entities by evaluating the impact of economic shocks on the network topology. This flexibility allows comparing the relative importance of factors for a given set of asset prices in a specific economic environment. While some of the connections are more persistent, other links are disabled and enabled according to the different economic or financial conditions.

We demonstrate these features by applying the framework to few different contexts: industry indices, blue chip stocks and financial institutions and analyze the dynamic relevance of Fama-French factors (Fama and French, 1993). We utilize appropriately designed network metrics to provide a summary of the characteristics of the network. We notice that these networks changing in time very frequently and that these patterns share commonalities for industry level and stock level returns.

The rest of the paper is organized as follows. In the next section, we describe the construction of MST financial networks with explanatory co-dependence features, as well as define network metrics suitable to measure network characteristics. In Section 3, we apply this framework to the different financial application contexts. We show how the network behaves when we change the window size and the correlation metrics in section 4. In section 5, we discuss the results and conclude.

2 Modeling Explanatory Dependencies

In this paper, we develop a framework that possesses explanatory capacity for temporal co-dependencies in asset price movements. As such, given the number of assets or firms under study, such a model can be quite complex in structure, with numerous inter-asset or interfirm linkages. Our model is minimalist in identifying the most significant interdependencies among asset prices and how these interdependencies change with time.

We wish to develop a dependency structure for N asset prices, $Y_i(t)$, $i \in \{1 \dots N\}$. Creating a minimal dependency structure, such as a minimum spanning tree, based on pairwise asset price correlation (Mantegna, 1999) fails to construct an explanatory indicator for these connections. We identify a set of explanatory factors, $X_j(t)$, $j \in \{1 \dots M\}$, to include in the development of the co-dependency structure. In this design, the asset prices and explanatory factors are identified as nodes, with each group of nodes forming a network layer.

The role for explanatory variables in our model is to verify how determinant they are in explaining asset price co-movement. Asset prices may move due to these explanatory variables, or may co-move due to their idiosyncratic, or firm-specific unique, characteristics. In order to elucidate this possible explanatory role, we split each asset price node into its systematic component $Y_i'(t)$, $i \in \{1 \dots N\}$, and idiosyncratic or unique component, $Z_i(t)$, $i \in \{1 \dots N\}$. This is accomplished by implementing the following factor model.

$$Y_i = \sum_{j=1}^{M} r_{ij} X_j + \sqrt{1 - (\sum_{j=1}^{M} r_{ij})^2} Z_i, \tag{1}$$

where each term of the factor model is described as follows:

- Y_i: return of asset i;

- r_{ij}: factor loading of i^{th} asset on j^{th} explanatory factor;

- X_j: explanatory factors (Note that X_j as a factor can assume negative values.);

- Z_i: idiosyncratic component of asset price evolution;

- $\sqrt{1 - (\sum_{j=1}^{M} r_{ij})^2}$: idiosyncratic factor loading.

The above factor model is implemented by applying ordinary least squares regression of the asset returns on the returns of the M explanatory factors. The residual of the regression is taken as the idiosyncratic portion of asset price movements not explained by the explanatory factors. This decomposition of asset price nodes results in $2N + M$ nodes in the network, where there are M explanatory factors, N asset price nodes and N idiosyncratic component of asset nodes. In this expanded network, the construction of the minimum spanning tree must be modified to serve its purpose. Note that, by selecting only the most relevant connections to each node, we filter out spurious links.

The co-movement between asset prices, explanatory factors and idiosyncratic component is measured using a dependence measure. For the purpose of this paper, we use four different dependence measures, Pearson ρ^P, Kendall ρ^K, Spearman ρ^S and weighted ρ^W.

$$\rho^P_{AB} = \frac{\sum_{i=1}^n (A_i - \bar{A}_i)(B_i - \bar{B}_i)}{\sqrt{\sum_{i=1}^n (A_i - \bar{A}_i)^2}\sqrt{\sum_{i=1}^n (B_i - \bar{B}_i)^2}}, \tag{2}$$

$$\rho^K_{AB} = \frac{1}{\binom{n}{2}} \sum_{1 \le p \le q \le n} sign((A_p - A_q)(B_p - B_q)) \tag{3}$$

$$\rho^S_{AB} = \frac{12}{n(n^2-1)} \sum_{i=1}^n \left(rank(A_i) - \frac{n+1}{2} \right) \left(rank(B_i) - \frac{n+1}{2} \right) \tag{4}$$

$$\rho^W_{AB} = \frac{\sum_{i=1}^n \alpha^{n-i} (A_i - \bar{A}_i)(B_i - \bar{B}_i)}{\sqrt{\sum_{i=1}^n \alpha^{n-i} (A_i - \bar{A}_i)^2}\sqrt{\sum_{i=1}^n \alpha^{n-i} (B_i - \bar{B}_i)}} \tag{5}$$

As such other measures for co-movement can be selected, however unless noted otherwise, we use Pearson correlation. Therefore, we measure intra-layer and inter-layer correlations ρ^{yx}_{ij}, between asset prices, explanatory factors ρ^x_{ij} and idiosyncratic components ρ^z_{ij}.

In order to develop a minimalist co-dependence structure, we adopt the minimum spanning tree (MST) approach widely studied in the literature Mantegna (1999), to identify the most significant intra-layer and inter-layer links of the network. Constructing an MST requires transforming the co-movement measure, in our case correlations, into a distance measure. A distance measure for inter-layer and intra-layer nodes is constructed as follows.

$$d^z(i,j) = \sqrt{1 - |\rho^z_{ij}|}, \tag{6}$$

$$d^x(i,j) = \sqrt{1 - |\rho^x_{ij}|}, \tag{7}$$

$$d^{yx}(i,j) = \sqrt{1 - |\rho^{yx}_{ij}|}, \tag{8}$$

We acknowledge the fact that negative correlation is just as much an indicator of co-dependence as positive correlation is, and follow Zheng et al. (2013) in using a distance metrics that takes into consideration the absolute value of the correlation. This is especially important for relationship between explanatory factors and asset prices. It can be shown that the above distance measure satisfies the properties of a norm. For instance,

1. $d^{(\cdot)}(i,j) = 0$ if and only if $i = j$,

2. $d^{(\cdot)}(i,j) = d^{(\cdot)}(j,i)$, and

3. $d^{(\cdot)}(i,j) \leq d^{(\cdot)}(i,k) + d^{(\cdot)}(k,j)$.

The distance measure, $d^{(\cdot)}(i,j)$, is used to determine the inter- and intra-layer links to construct the minimal spanning tree connecting the N asset prices, N idiosyncratic components and M explanatory factors using $2N + M - 1$ links. It should be noted that the higher the correlation between nodes, positive or negative, the shorter the distance between the nodes by this chosen distance measure.

The modified minimum spanning tree is constructed as follows. We construct the distance matrix, $[d^{(\cdot)}(i,j)]$, for the entire set of nodes as the first step. All the asset nodes, N of them, are connected to their corresponding idiosyncratic component nodes. This forms N two-node subnetworks of the assets. The complete network is formed by the connections between these subnetworks via the idiosyncratic nodes or by connections of these subnetworks with the explanatory factor nodes via the asset node.

The identification of the rest of the links to construct a minimal spanning tree follows a greedy sequential approach of discovering the links. The upper-diagonal entries of the distance matrix constructed for idiosyncratic node interaction and asset node-explanatory variable interaction is sorted in an increasing order. These sorted values are used to identify links of a minimal spanning tree, picking the most significant links that minimally connect the entire set of nodes. From the lowest to the highest values of pair-wise distance measures, links are constructed sequentially for each node so that the node is connected to the network through the closest (or shortest distance measure) node, until all the nodes are connected.

The layered minimum spanning tree for the explanatory co-dependence structure among asset prices can indeed change with time. This structural evolution can be instructive at the overall network level, as well as at a single asset or firm level. Moreover, studying temporal characteristic of the co-dependence structure can provide insight regarding the role of explanatory factors and impact of economic shocks. Therefore, we are interested in observing the change in the minimum spanning tree network constructed above through time.

The temporal evolution of the minimum spanning tree is constructed by rolling the window of observations while constructing the layered network. Each window of observations it taken to have T data points, for instance, we use a 120 day or approximately 6 months window in our study. At each point in time t, there is a set of $Y_i(t)$ and $X_j(t)$ observations, from which the idiosyncratic terms, $Z_i(t)$, are extracted. With all the nodes defined, $\rho(t)$ and $d_{ij}(t)$ are computed for time t as a function of the T observations. This generates a unique layered network for each time t. We then move the rolling-window by a time-increment, Δt, and re-construct the network. This helps us evaluate the topological variations in co-dependence structure for the asset prices over time.

In our model development thus far, we have described the construction of static and dynamically-evolving two-layered explanatory co-dependence network structure for asset prices. In order to analyze the properties of this network, we need to define appropriately constructed network measures. These network measures must be developed keeping in view the two-layered structure and three types of nodes, explanatory factor nodes, X, systematic component of asset prices nodes, Y and idiosyncratic component of asset price nodes, Z. There are the inter-layer links, which perform the primary explanatory role between asset prices and explanatory variables, and intra-layer links that indicate how idiosyncratic connections among assets occur or how explanatory factors are linked together.

The analysis of the network requires understanding the nature of connectivity between nodes in terms of node degrees and link sizes. Node degree deg_i is the number of links connecting a given node to the rest of the network. Higher node degrees in a network indicate clustering, i.e., a few nodes assuming significant, central position. The link size l_{ij} is the distance between two connected nodes i and j. Our distance measure, $d_{ij}(t)$, is a monotonically decreasing function of the absolute value of correlations, $\rho(t)$, thus smaller link sizes indicate stronger connection between entities.

A measure of network's overall strength of connectivity is given by the total link size for all the links of the network. The average link size normalizes this measure. We begin the analysis with the overall strength of connectivity of the network by calculating the average link size, irrespective of the layer the links belong to. Therefore, we define

$$\bar{l} = \frac{\sum_{k=1}^{2N+M-1} l_k}{2N+M-1}, \tag{9}$$

where we note that the minimum spanning network has $2N+M$ nodes and that l_k is the kth of the $2N+M-1$ connected l_{ij} links . The average size of the links shows the average strength and the average size of the network. We similarly define and analyze the average strength of connectivity for intra-layer links and inter-layer links using the link sizes of: inter-layer connections l_{xy}, and the average link size intra-layer explanatory l_{xx} links and idiosyncratic l_{zz} links.

In the temporal evolution of the minimum spanning tree, at any point in time, a drop in the average link size of inter-layer connections $\bar{l}_{xy}$ indicates connections between explanatory factors and asset prices getting stronger. This may happen due to the existing inter-layer connections becoming stronger or because the weaker inter-layer connections getting disabled. The average link size of intra-layer connections among explanatory factors, $\bar{l}_{xx}$, indicates how systemically close the explanatory factors are moving. A higher value of $\bar{l}_{xx}$ indicates that the connected factors are connected only weakly. Finally, a low value of the average link size of intra-layer idiosyncratic links $\bar{l}_{zz}$ indicates that asset prices are strongly connected due to their idiosyncratic characteristics.

In a minimum spanning tree, although the number of links of the network is constant, $2N + M - 1$ in our case, the dispersion of node degree values is an indicator of the topology of the network. We use two metrics to represent this dispersion, standard deviation of node degrees and maximum node degree. Standard deviation of node degrees as our dispersion measure is defined as follows,

$$\sigma_{Deg} = \sqrt{\frac{\sum_{i=1}^{2N+M}(deg_i - \overline{deg})^2}{2N + M}}, \tag{10}$$

where $\overline{deg}$ is the average node degree of all the nodes.

The node degree of each of the $2N + M$ nodes of the network is greater than or equal to 1, and is at most $2N + M - 1$. The standard deviation of node degrees holds information regarding the structure of network topology. Networks highly concentrated around a couple of nodes will tend to have a high standard deviation of node degrees, while in sparse decentralized networks, nodes will tend to have similar node degrees. Thus, chain-like networks will have a lower standard deviation of node degrees and star-like networks will have a greater standard deviation of node degrees. Standard deviation of node degrees can also be defined specifically for each type of node, explanatory factor, idiosyncratic or asset price node.

An alternative measure of dispersion is the highest node degree of the network. This measure represents the number of links of the most connected node and provides insight on the important role the most connected node plays in the network. Higher maximum node degree is associated with star-like networks. where one node has higher importance in connecting the other nodes to the network.

3 Empirical Analysis

We apply the framework of developing an explanatory minimal dependence structure in asset prices to a few contexts and discuss the results. We first study the overall US economy, since the 1970s, by examining the relationship between industry portfolios. Then, we focus on blue chip stocks since the 1990s.

The framework requires selecting factors considered suitable for developing the explanatory dependence structure among asset prices. Since the development of the Capital Asset Pricing Model (Sharpe, 1964), researchers have studied asset prices from the perspective of their linear relationship with explanatory factors. In Fama and French (1993), the authors show that apparent anomalies of the CAPM can be explained with the inclusion of new factors giving rise to multi-factor literature. The authors identify both firm size and book to market value of equity ratio to have a role in determining the cross-section of average equity returns. The explanatory layer of the networks is constructed using these three common Fama-French risk factors for equity: market return, size factor (SMB) , and value (HML) factors. Data and detailed explanation for these factors can be obtained from the Kenneth

French Data Library [1].

3.1 Industry Indices

In Fama and French (1997), the authors note that the three-factor risk loadings for industries are not constant and a full sample estimation could generate inaccurate estimations of the dependence structure. Our framework can help understand how these risk-loadings are changing in time.

We use an updated data set from Kenneth French Data library. As in Fama and French (1997), each stock from NYSE, AMEX and NASDAQ is assigned to an industry portfolio based on its SIC code. The updated version of this data set includes 49 industries. We start our analysis in 1970 and observe daily closing prices until 2014. For each day in the sample, we construct the unlayered minimum spanning tree among all nodes and the layered minimum spanning tree using the explanatory factors, as discussed in section 2, using a rolling window of 120 business days.

The resulting network characteristics are summarized in plots of Figure 1. In the top graph of the figure, we indicate a comparison of average distance between nodes in the flat or unlayered network versus in the explanatory layered network. In the flat unlayered network, all nodes are identical and represent one of the 49 industries, however in the explanatory layered network, there are three types of nodes, the explanatory factor nodes, idiosyncratic nodes and pure nodes, represented by returns of the industry portfolios. In this plot, the average distance does not discriminate the node type from which the distance is measured. Since the total number of connections in a minimum spanning tree is a constant, the average

[1] Available at $http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html$

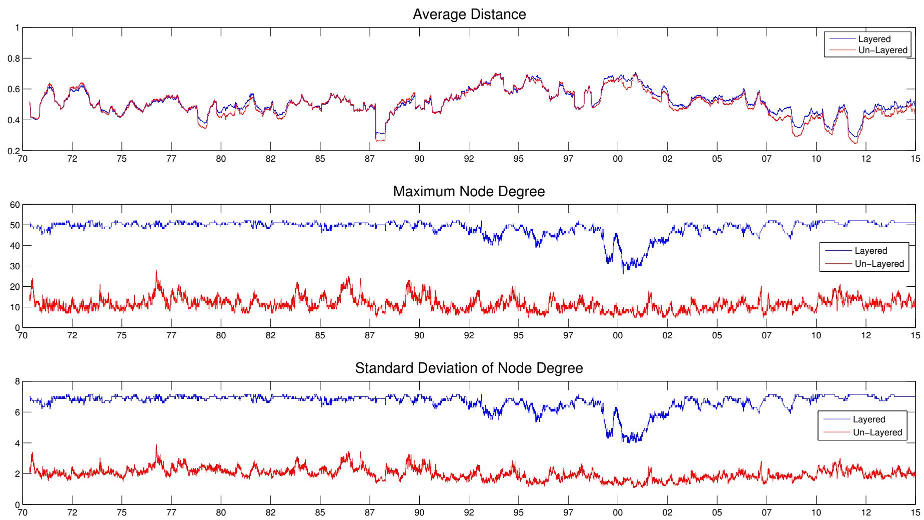

Figure 1: Comparing layered and unlayered network in terms of average distance between nodes (top graph), highest node degree (middle graph) and standard deviation of node degree (bottom graph) for 49 industries

distance between nodes of the networks provides insight on the overall strength of connections between entities of the network.

As seen in the average distance plot of figure 1, there is significant variability of this metric for both networks. The average distance for both types of networks move together at all times, suggesting that the information provided by an unlayered network is not lost in this new framework. Over time, the average distance reduces. These stronger of connections show that the dependence measure is higher across the system and thus co-movement is rising. Also, there are sudden reductions of the distance for small periods of time that seem to coincide with financial crises, such as 1987, 1997, 2008 and 2011. This metric, however, does not provide information if the shape of the network is changing and whether the network is becoming tighter due to firm specific characteristics or to common explanatory factors.

The second and third plots of the figure provide information on the shape of the networks. The maximum node degree shows the importance of the most connected node to the network, while the standard deviation of the node degree sheds light on the dispersion of node degrees across the network. By design, these two metrics bear a very strong relationship. These dispersion metrics change very frequently for the unlayered network, but the layered network is more stable and abrupt shape changes don't happen as often. We notice that the layered network is very star-like, i.e. few nodes concentrate most of the links until the the end of the 1990s, when its shape changes dramatically until 2002. After that, the network changes back to its original star-like shape.

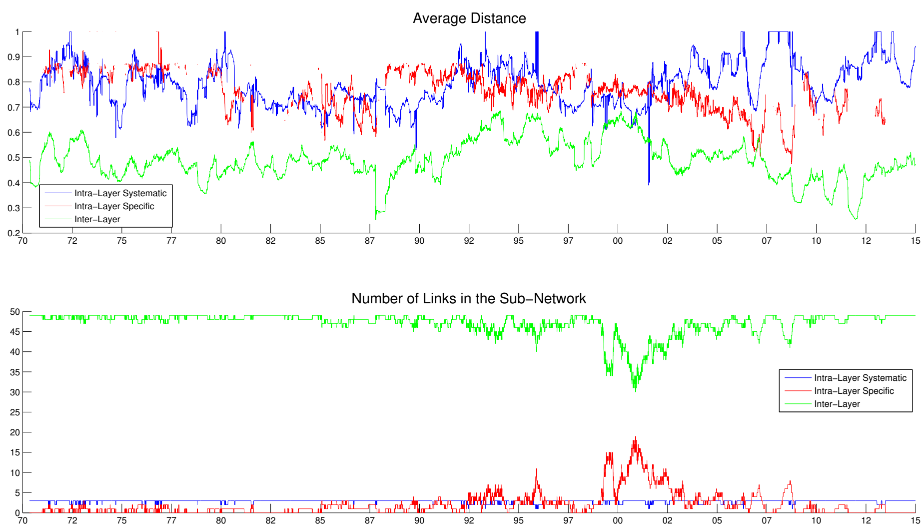

Figure 2: Network characteristics by node type in explanatory layered network, in terms of average distance between inter- and intra-layer nodes (top graph), number of connections intra- and inter-layer (bottom graph) for 49 industries

A natural element to investigate in the shape of the layered network is the relationship between the different types of nodes. In Figure 2, we present inter- and intra-layer network properties. In the top graph of the figure, average distance for only the inter- and intra-layer links are displayed. The inter-layer average distance indicates the strength of explanatory connections for the financial institutions over time, while intra-layer, especially within the idiosyncratic layer, indicates unique linkages between financial institutions. In the lower graph of the figure, we display the number of inter- and intra-layer links. As such for $2N + M$ nodes in the network, the total number of links in a minimum spanning tree is $2N + M - 1$. Therefore, the point of attention here is how many of these $2N + M - 1$ links are inter-layer and intra-layer, respectively.

The top graph of Figure 2 shows that the inter-layer and idiosyncratic intra-layer average distance displays significant modulation through time. When compared with plot of Figure 1, the inter-layer average distance seen in this plot shows the contribution of strengthening of connections between industries arising from explanatory factors. In the bottom plot, we see that the change in shape of the late 1990s and early 2000s, is driven by the increased importance of idiosyncratic connections. We also note that the presence of these connections is not necessarily due to augmented co-movement, but rather to lack of explanatory power from the factors.

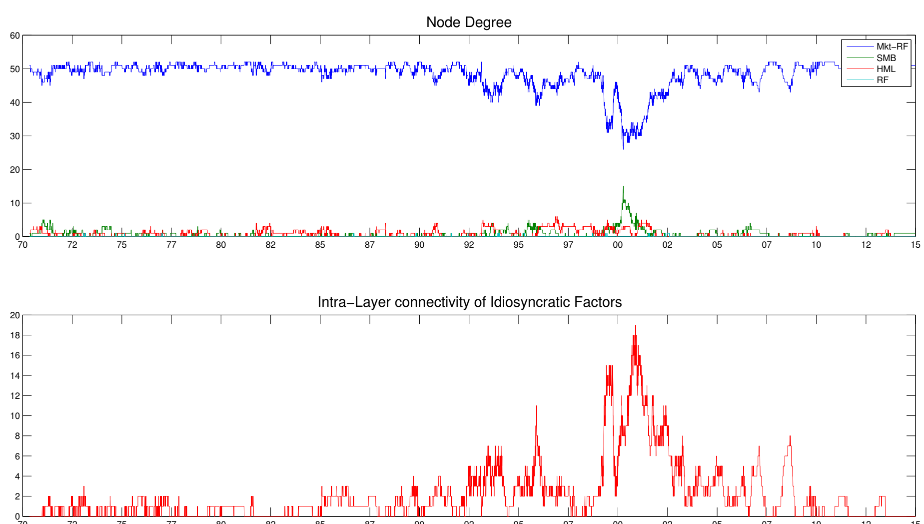

Figure 3: Node characteristics in explanatory layered network, in terms of degree of explanatory factor nodes (top graph), number of connections between idiosyncratic nodes (bottom graph) for 49 industries

In Figure 3, we show the node degree for each of the explanatory factors. As expected, the market factor is prevalent during the whole sample period. However, during the period between 1999 and 2002, the importance of this factor is reduced and connections involving the SMB factor and among idiosyncratic nodes are enabled. We use the insights obtained thus far from the explanatory layer network characteristics to select six specific time points for which we display the actual layered network.

In order to aid visualization of a network with a large number of nodes, the layered network for these time points is shown in a carefully constructed format in Figure 4. The systematic or explanatory factors are placed in an arc as the top layer, the core firm nodes form the next layer, while the idiosyncratic nodes are organized on an arc in the bottom layer. This facilitates displaying connection of core firm nodes with explanatory factors, and intra-layer connections among explanatory factors and idiosyncratic nodes, respectively. Links are colored according to the sign of their dependence metric, blue links represent positive co-movements and red links represent negative co-movements.

In 1987, only two industries are not directly linked to the market risk factor, they were gold and mines. Besides being linked to each other, these industries are linked to the rest of the network by the the idiosyncratic link between mines and autos. It is remarkable how the explanatory power of the factors change during the 2 years around the turn of the century. On January of 2000, 13 assets are not linked to the market risk factor. In general, they were consumer non durables (agriculture, food soda, beer, tobacco products, books), transportation (ships), mining and minerals (non-metallic and industrial metal mining, gold), utilities, software and banks. We notice some negative links between factors and nodes, n particular, the SMB is directly linked to industries such as food products, beer and liquor,

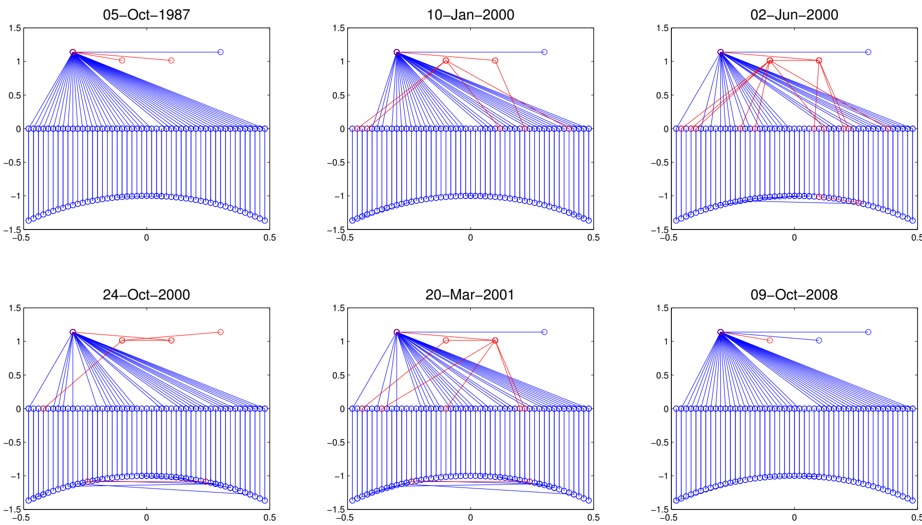

Figure 4: Display of networks for six time points along the observation period. The explanatory factors are (from right to left) RM-RF,SMB, HML, TERM.

utilities and the HML was directly linked to the software industry. On June of 2000, the 21 nodes are not connected to the market risk factor. Industries disconnecting from the market risk factor are close to those that were already disconnected, but many were still connected to the SMB and HML. By October of 2000, these industries remain disconnected from the market risk factor, but become connected among them by idiosyncratic links. We also note that during the 2008 financial crisis, only 6 industries are not directly linked to the market risk factor. They are all connected among them, on a subnetwork of mininng and energy industries, that connects to the rest of the network by the steel industry.

3.2 Blue Chip Stocks

Having identified that this network of industry indices has some patterns, we move on to investigate whether the same pattern is present on a group of blue chip stocks. For this analysis, we select all stocks that were part of the Dow Jones Industrial Average Index since the 1990s. We drop from the sample three firms that had their IPO after 1990 and do not have data for the whole period.

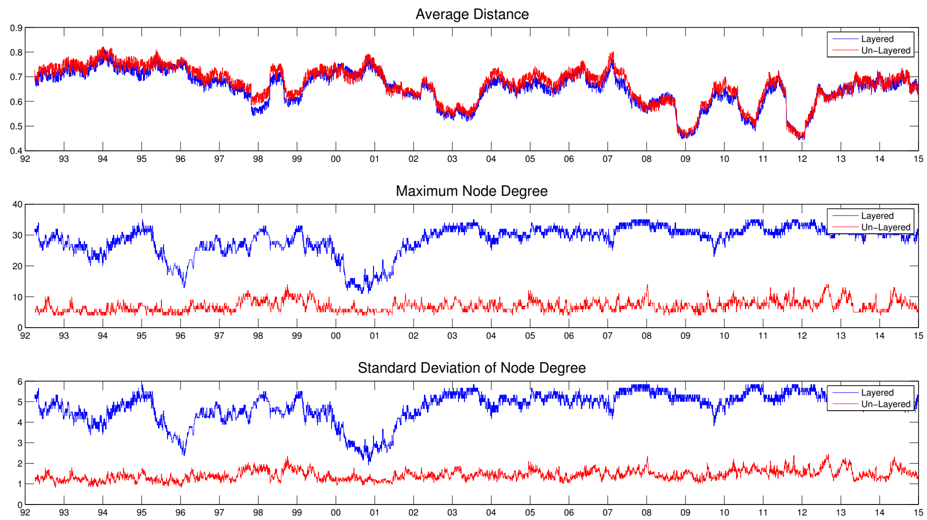

Figure 5: Comparing layered and unlayered network in terms of average distance between nodes (top graph), highest node degree (middle graph) and standard deviation of node degree (bottom graph) for 33 Blue Chip Stocks

We see a strong relationship between the network metrics shown in Figures 1 and 5. When we examine the average link distance, peaks and troughs are coincident suggesting that the dependence between industries strengthening of the industry level network happens at the same time as the network on the firm level. The same happens for the shape of the network, in particular the large decrease of the maximum node degree around in 1995 and around the time of the dot com bubble bust.

In fact, when we observe the evolution of the node degree of factors, we notice that during the two periods aforementioned, the market factor loses importance to the SMB. In these two particular intervals, there are also high numbers of intra-layer idiosyncratic connections hinting to an apparent reduction on the importance of the Fama-French factors in explaining returns. Another important and relevant finding is that, after the financial crisis, the market

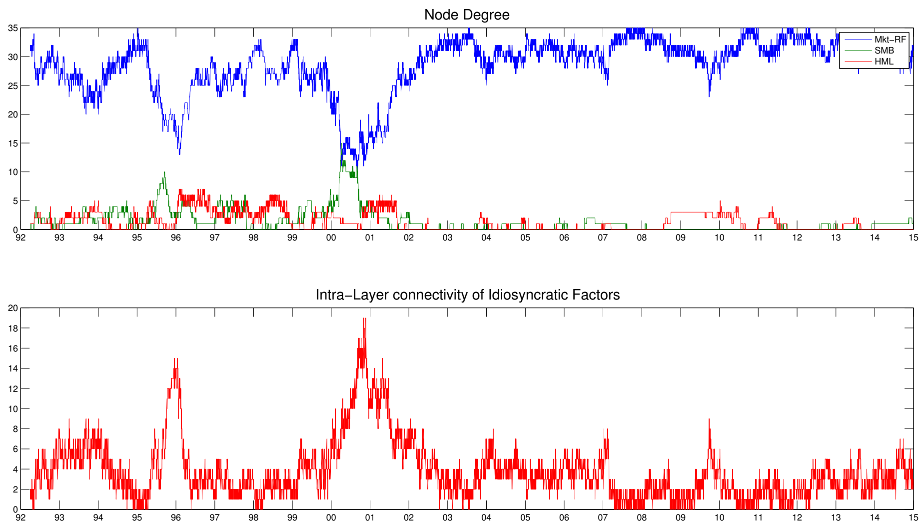

Figure 6: Node characteristics in explanatory layered network, in terms of degree of explanatory factor nodes (top graph), number of connections between idiosyncratic nodes (bottom graph) for 33 blue chip stocks

factor appear as the sole connected factor. This evidence is consistent with the empirical findings that co-movement increases during high volatility periods Boyer et al. (2006).

4 Robustness of the Network

A natural question that arises from the analysis of the results on section 3 concerns the robustness of the network and its metrics under different assumptions, i.e., what happens when the window size or the co-movement measure changes. In this section, we show a couple of different ways of selecting and observing the network. As expected, much like correlation analysis, the information filtered by the network will vary according to the parameters used.

There is a trade-off between information and stability when choosing the window size of estimation. Shorter windows will yield more rapid updates of the network, while longer windows will provide more stable networks. For applications that require higher frequency updates, however, a minimum number of data points is necessary to provide consistent estimates. For instance, while the unlayered network has $\frac{n(n-1)}{2} \sim \frac{n^2}{2}$ input parameters, the layered network has $\frac{(f+2n)(f+2n-1)}{2} \sim \frac{(f+2n)^2}{2}$. Since there are NT observations of asset returns and FT observations of factor returns, this means that we need $T >> \frac{n}{2}$ for the unlayered network to be estimated and $T >> \frac{f+4n}{2}$, which means that, for our cases, 120 is the most appropriate window size.

Another concern could come from the convenience of overlapping rolling windows, once by rolling only one day at a time, our network carries each observation for the duration of the window. We use the example of 49 industries and 3 factors to explore these two questions. In figure 7, we plot non-overlapping windows of 60, 120 and 250 days in addition to our base case.

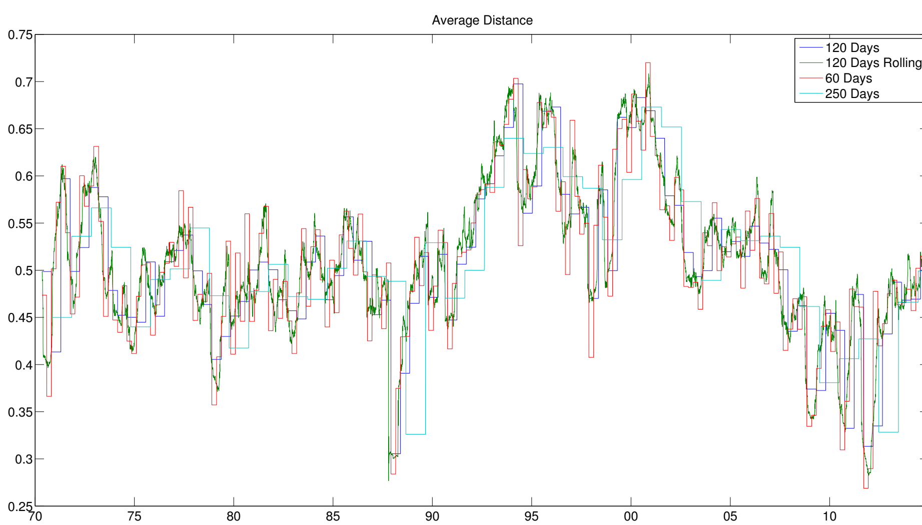

Figure 7: Display of average distance of networks based on 60, 120 and 250 business days correlation of returns of industries and factors

Indeed, in line with Bonanno et al. (2001), we find that the choice of window size matters for the networks. Larger window sizes are more stable, however, they can hide changes in the network topology. For example, in 2008, during the international financial crisis, the average distance of the network decreased dramatically for windows of 60 and 120 days and later increased again as the most acute period of the crisis was over. However, the 250 day window did not capture this change. On the other hand, in other periods, shorter intervals may generate excessive jumps on the network size, such was the case in the crises of the second half of the 1990s, when networks based on 60 days windows showed a greater variance on their distance.

We rely on the same example to investigate the robustness of the network for different comovement measures. In figure 8 we show the average distance of the network for four different comovement metrics: Pearson, Kendall, Spearman and weighted. We note that overall the metrics suggest that the size of network does not change dramatically when the comovement measure is changed. An exception is made to the weighted correlation. Not only they exhibit greater variance, which was expected given that more this metric gives larger weight for more recent observations and, thus, a shorter window prevail, but also the average average distance is usually smaller than in the other networks.

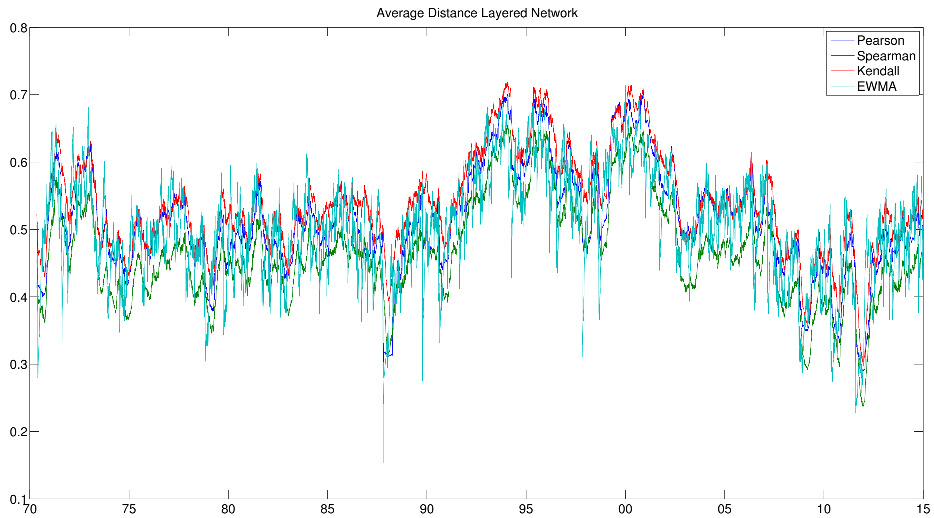

Figure 8: Display of average distance of networks based on 120 business days and different comovement measures

In figure 9, we see how the topology changes over time for the networks based on the different comovement measures. Similarly, in this case, Pearson, Kendal and Spearman correlation yield corresponding metrics, suggesting that the network is about the same for them. The weighted correlation network, however, is less stable, changing more frequently and more drastically. Additionally, a reduced standard deviation suggests that more idiosyncratic connections are enabled over time in this network.

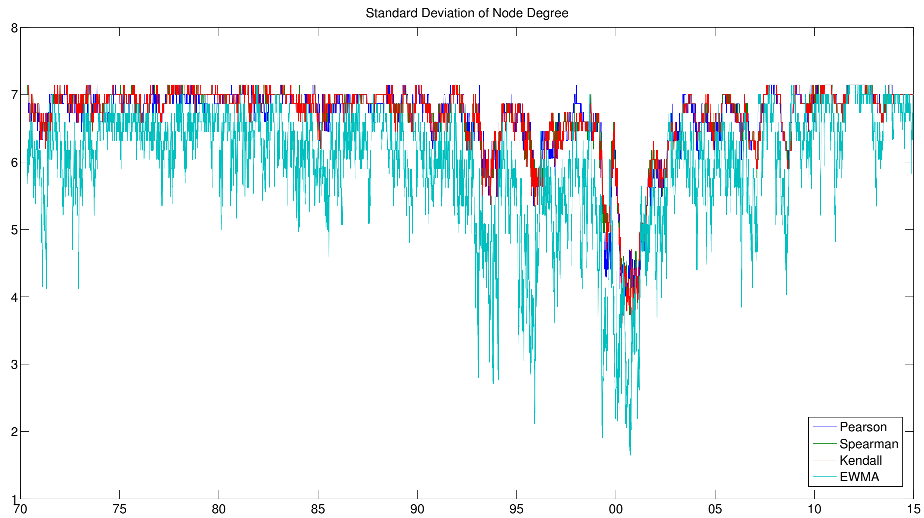

Figure 9: Display of standard deviation of node degree of networks based on 120 business days and different comovement measures.

5 Discussion and Conclusion

In this work, we provide a framework for modeling co-dependencies in asset price movements, by creating a layered minimum spanning tree structure. This new methodology adds to the previous literature of financial networks by allowing for explanatory factors to contribute in describing connections between asset price nodes. The temporal changes in topology of this type of network has also not been studied before.

The network is constructed following a greedy sequential approach of discovery of the links. Asset prices and explanatory factors are nodes from different layers of the minimalist interdependence structure. Connections between asset prices are allowed to occur through the explanatory variables or firm-specific idiosyncratic characteristics. We evaluate the role of these connections by investigating changes in topology of the network and strength of the enabled connections.

Another feature of this methodology is the flexibility in the choice of the explanatory factors to suit the specific objectives of the study and the context of constructing minimalist interdependency connections between nodes. The methodology permits comparisons of the relative importance of factors for a given set of asset prices. This flexibility may serve the purpose of decomposing the importance of different factors by adding or removing them from the subset of explanatory factors.

The temporal characteristic of the network enhances our understanding of the links between the entities by evaluating the impact of economic shocks in the network topology. While some of the connections are more persistent, other links are disabled and enabled according to the different economic or financial conditions. This opens space for other analyses regarding the persistence of these links and what is behind these changes in strength of the connections.

In this paper we presented 3 cases for the application of the methodology. We start by looking at industry level indices for the whole market, then we focus on large individual stocks and later we focus on stocks from a single industry. We observe that the methodology is helpful in understanding the changes between the interdependencies among entities in the data-set. In particular, we were able to see dramatic changes in the topology of the network in crisis periods of the financial system. The variety of changes in the topology and strength of connections we observed in these cases deserve further exploration.

The cases shown here also highlight the flexibility of this framework in analyzing financially related phenomena. The methodology is agnostic to the underlying data and to the choice of dependency structure. This may serve purposes of risk management and portfolio allocation in contexts other than the analysis of market risk provided here.

References

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Aste, T., Di Matteo, T., Tumminello, M., and Mantegna, R. N. (2005). Correlation filtering in financial time series. In *SPIE Third International Symposium on Fluctuations and Noise*, pages 100–109. International Society for Optics and Photonics.

Barro, D. and Basso, A. (2010). Credit contagion in a network of firms with spatial interaction. *European Journal of Operational Research*, 205(2):459–468.

Bonanno, G., Caldarelli, G., Lillo, F., and Mantegna, R. N. (2003). Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130.

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

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

Boyer, B. H., Kumagai, T., and Yuan, K. (2006). How do crises spread? evidence from accessible and inaccessible stock indices. *The Journal of Finance*, 61(2):957–1003.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). Modeling international financial returns with a multivariate regime-switching copula. *Journal of Financial Econometrics*, 7(4):437–480.

Das, S. R. and Uppal, R. (2004). Systemic risk and international portfolio choice. *The Journal of Finance*, 59(6):2809–2834.

De Masi, G. and Gallegati, M. (2012). Bank–firms topology in italy. *Empirical Economics*, 43(2):851–866.

Engle, R. and Figlewski, S. (2015). Modeling the dynamics of correlations among implied volatilities. *Review of Finance*, 19(3):991–1018.

Engle, R. F., Weidman, S., and Reed, J. (2010). *Technical Capabilities Necessary for Regulation of Systemic Financial Risk: Summary of a Workshop*. National Academies Press.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56.

Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of financial economics*, 43(2):153–193.

Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance*, 57(5):2223–2261.

Gai, P., Haldane, A., and Kapadia, S. (2011). Complexity, concentration and contagion. *Journal of Monetary Economics*, 58(5):453–470.

Huang, W.-Q., Zhuang, X.-T., and Yao, S. (2009). A network analysis of the chinese stock market. *Physica A: Statistical Mechanics and its Applications*, 388(14):2956–2964.

Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., and Ben-Jacob, E. (2010). Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market. *PloS one*, 5(12):e15032.

Kohonen, A. (2013). On detection of volatility spillovers in overlapping stock markets. *Journal of Empirical Finance*, 22:140–158.

Kole, E., Koedijk, K., and Verbeek, M. (2007). Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8):2405–2423.

Kwapień, J., Gworek, S., Drożdż, S., and Górski, A. (2009). Analysis of a network structure of the foreign currency exchange market. *Journal of Economic Interaction and Coordination*, 4(1):55–72.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets. *The Journal of Finance*, 56(2):649–676.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *The European Physical Journal B-Condensed Matter and Complex Systems*, 11(1):193–197.

Minoiu, C. and Reyes, J. A. (2013). A network analysis of global banking: 1978–2010. *Journal of Financial Stability*, 9(2):168–184.

Mizuno, T., Takayasu, H., and Takayasu, M. (2006). Correlation networks among currencies. *Physica A: Statistical Mechanics and its Applications*, 364:336–342.

Naylor, M. J., Rose, L. C., and Moyle, B. J. (2007). Topology of foreign exchange markets using hierarchical structure methods. *Physica A: Statistical Mechanics and its Applications*, 382(1):199–208.

Onnela, J.-P., Chakraborti, A., Kaski, K., Kertesz, J., and Kanto, A. (2003). Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5):056110.

Papadimitriou, T., Gogas, P., and Tabak, B. M. (2013). Complex networks and banking systems supervision. *Physica A: Statistical Mechanics and its Applications*, 392(19):4429–4434.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482.

Tabak, B. M., Luduvice, A. V. D., and Cajueiro, D. O. (2011). Modeling default probabilities: The case of brazil. *Journal of International Financial Markets, Institutions and Money*, 21(4):513–534.

Tse, C. K., Liu, J., and Lau, F. (2010). A network perspective of the stock market. *Journal of Empirical Finance*, 17(4):659–667.

Tumminello, M., Lillo, F., and Mantegna, R. N. (2010). Correlation, hierarchies, and networks in financial markets. *Journal of Economic Behavior & Organization*, 75(1):40–58.

Vandewalle, N., Brisbois, F., Tordoir, X., et al. (2001). Non-random topology of stock markets. *Quantitative Finance*, 1(3):372–374.

Zheng, Z., Yamasaki, K., Tenenbaum, J. N., and Stanley, H. E. (2013). Carbon-dioxide emissions trading and hierarchical structure in worldwide finance and commodities markets. *Physical Review E*, 87(1):012814.

6 Appendix

Details of the entities included in the analysis of Section 3 are provided here. The explanatory factors we use are given in the table below.

Table 1: List Of Explanatory Factors

#	Ticker	Name
1	RMRF	Market Return - Risk Free Return
2	SMB	Small minus Big
3	HML	High minus Low

Table 2: List of Industries

#	Description		#	Description
1	Agriculture		26	Defense
2	Food Products		27	Precious Metals
3	Candy & Soda		28	Non-Metallic and Industrial Metal Mining
4	Beer & Liquor		29	Coal
5	Tobacco Products		30	Petroleum and Natural Gas
6	Recreation		31	Utilities
7	Entertainment		32	Communication
8	Printing and Publishing		33	Personal Services
9	Consumer Goods		34	Business Services
10	Apparel		35	Computers
11	Healthcare		36	Computer Software
12	Medical Equipment		37	Electronic Equipment
13	Pharmaceutical Products		38	Measuring and Control Equipment
14	Chemicals		39	Business Supplies
15	Rubber and Plastic Products		40	Shipping Containers
16	Textiles		41	Transportation
17	Construction Materials		42	Wholesale
18	Construction		43	Retail
19	Steel Works Etc		44	Restaurants, Hotels, Motels
20	Fabricated Products		45	Banking
21	Machinery		46	Insurance
22	Electrical Equipment		47	Real Estate
23	Automobiles and Trucks		48	Trading
24	Aircraft		49	Almost Nothing
25	Shipbuilding, Railroad Equipment			

Table 3: List of Blue Chip Stocks

#	**Stock**	**Industry**
1	General Electric	Industrial
2	3M	Industrial
3	United Technologies	Industrial
4	Honeywell international inc	Industrial
5	Home Depot	Consumer, Cyclical
6	Mcdonald's	Consumer, Cyclical
7	Nike	Consumer, Cyclical
8	Walmart	Consumer, Cyclical
9	Motors Liquidation	Consumer, Cyclical
10	Johnson & Johnson	Consumer, Non-cyclical
11	Coca-Cola	Consumer, Non-cyclical
12	Merck & Co.	Consumer, Non-cyclical
13	Pfizer	Consumer, Non-cyclical
14	Procter & Gamble	Consumer, Non-cyclical
15	United Health Group	Consumer, Non-cyclical
16	Altria Group	Consumer, Non-cyclical
17	IBM	Technology
18	Intel	Technology
19	Microsoft	Technology
20	Hewlett-Packard	Technology
21	Eastman Kodak co	Technology
22	JPMorgan Chase	Financial
23	Travelers	Financial
24	Bank of America	Financial
25	Citigroup inc	Financial
26	Verizon Communications	Communications
27	AT&T	Communications
28	Exxon Mobil	Energy
29	Alcoa	Basic Materials
30	International Paper	Basic Materials