Topological properties of stock market networks: The case of Brazil

Benjamin M. Tabak\textsuperscript{a,b,*}, Thiago R. Serra\textsuperscript{c}, Daniel O. Cajueiro\textsuperscript{c,d}

\textsuperscript{a} Banco Central do Brasil, Brazil
\textsuperscript{b} Graduate Program in Economics – Universidade Católica de Brasília, Brazil
\textsuperscript{c} Department of Economics – Universidade de Brasília, Brazil
\textsuperscript{d} National Institute of Science and Technology for Complex Systems, Brazil

\textbf{ARTICLE INFO}

\textbf{Article history:}
Received 22 July 2009
Received in revised form 18 February 2010
Available online 20 April 2010

\textbf{Keywords:}
Stock market network
Emerging markets
Econophysics
Minimal spanning tree

\textbf{ABSTRACT}

This paper investigates the topological properties of the Brazilian stock market networks. We build the minimum spanning tree, which is based on the concept of ultrametricity, using the correlation matrix for a variety of stocks of different sectors. Our results suggest that stocks tend to cluster by sector. We employ a dynamic approach using complex network measures and find that the relative importance of different sectors within the network varies. The financial, energy and material sectors are the most important within the network.

\copyright 2010 Elsevier B.V. All rights reserved.

\section{1. Introduction}

The analysis of stock markets is crucial for the development and design of investment strategies. Furthermore, understanding how stock markets evolve over time is important and useful. Therefore, describing patterns within stock markets has become a popular research agenda in the past years. The development of network theory has allowed to unveil these patterns in a very simple and elucidative way \cite{1–7}. We extend this analysis to the Brazilian equity sector.

The Brazilian equity market is an interesting case study for a number of reasons. First, it is one of the most important in Latin America and of emerging markets. Second, it has an important number of companies that have been involved in mergers and acquisitions worldwide and is becoming more integrated with international equity markets. Finally, some sectors have been targets of international companies such as telecommunications and understanding their role in the network is important.

This paper focuses on the Brazilian equity sector. The main objective is to characterize the topology and taxonomy of equity market networks. We also study the dynamic structure of the Brazilian stock market network, which provides relevant information regarding the importance of each sector. We calculate $\beta$-dominance, clustering, closeness and graph centrality, disparity and strength for each stock and evaluate how specific sectors behave over time. This provides information regarding the heterogeneity of the network and which sectors are more relevant. Our findings suggest that some sectors are more important within the network, due to a tighter connection with other sectors, which has important implications for the design of portfolio strategies. Furthermore, empirical results imply that the relevance of some sectors may increase in specific periods, which suggests that portfolio strategies that employ complex network tools should reassess their allocation decisions over time.

The remainder of the paper is structured as follows. Next section discusses the literature review for financial market networks, while Section 3 introduces the methodology and the sampling procedures. Section 4 shows the data and Section 5 presents empirical results. Finally, Section 6 provides some final considerations.

* Corresponding author at: Banco Central do Brasil, Brazil.
E-mail address: benjamin.tabak@bcb.gov.br (B.M. Tabak).

0378-4371/$ – see front matter \copyright 2010 Elsevier B.V. All rights reserved.
2. Brief literature review

Since the seminal works of Refs. [1,8] a large literature has provided overwhelming evidence that stock markets behave as complex systems. Furthermore, there is ample evidence that it is possible to study the topological properties of financial networks, which have an associated meaningful economic taxonomy [2,3,9–17,6,7,5,4,18,19]. In Ref. [20] one may find an interesting and new review of this literature. Overall this literature has suggested that it is possible to uncover the formation of clusters within stock market networks and that it is possible to define a taxonomy for the stocks within the network. These results are important and have many implications for the design of portfolio and risk management models.

The use of complex systems tools to analyze financial markets is important for a variety of reasons. Several papers have shown that the use of these tools may help in the design of portfolio strategies and risk assessment [21]. Besides, the characterization of the network topology may provide insights in the process of price formation.

Recent research has provided some evidence that financial networks evolve over time [11]. Nonetheless, so far there is not clear evidence on whether the parameters that characterize financial networks change over time. Only very few papers have studied this issue for emerging markets. Therefore, this paper seeks to contribute to this discussion. We follow this literature and contribute by characterizing the Brazilian stock market network and studying its changes over time.

3. Stock networks

We use the cross-correlation in stock price changes from January 7, 2000 to February 29, 2008.\(^1\) The dataset is given by weekly closing prices in terms of Brazilian Real, and the correlation matrix should represent short-term trends. From the log-return of the stock prices, which is defined as \(S_i(t) = \ln(Y_i(t)) - \ln(Y_i(t - 1))\), where \(Y_i(t)\) is the stock price of company \(i\) at time \(t\), we could calculate the cross-correlation function as:

\[
\rho_{ij} = \frac{\langle S_i \cdot S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\langle S_i^2 \rangle - \langle S_i \rangle^2}\langle S_j^2 \rangle - \langle S_j \rangle^2}}
\]

where \(\langle S_i \rangle\) represents the statistical average of \(S_i\), for a given time period [8] and the cross-correlations range from \(-1\) to \(1\).\(^2\)

To construct the MST, it is required a variable that can work as distance, satisfying the three axioms of Euclidean distance:

- (i) \(d_{ij} = 0\) if and only if \(i = j\),
- (ii) \(d_{ij} = d_{ji}\),
- (iii) \(d_{ij} \leq d_{ik} + d_{kj}\).

To build the stock prices network we employ the metric distance, proposed by Ref. [1], \(d_{ij} = \sqrt{2(1 - \rho_{ij})}\), where \(\rho_{ij}\) is the correlation between changes in stocks \(i\) and \(j\). The MST is the spanning tree of the shortest length using Prim’s algorithm of the \(d_{ij}\) and is a graph without cycles connecting all nodes with links. Prim’s algorithm is an algorithm in graph theory that finds a minimum spanning tree for a connected weighted graph. This means it finds a subset of the edges that forms a tree that includes every vertex, where the total weight of all the edges in the tree is minimized. If the graph is not connected, then it will only find a minimum spanning tree for one of the connected components.

The distance \(d_{*ij}\) between two successive stocks \(i\) and \(j\), is called subdominant ultrametric distance and a space connected by these distances provides a topological space that has associated an unique indexed hierarchy. The distance \(d_{*ij}\) satisfies the above axioms of Euclidean distance and also the following strong ultrametric inequality:

\[
d_{ij} \leq \max[d_{ik}, d_{kj}].
\]

To construct the hierarchical tree and a better interpretation of it we used the subdominant ultrametric distance \(d_{*ij}\) in the single-linkage clustering method. In this method, we use the distance between the closest pair of nodes \(i\) and \(j\).

The MST is the correlation based network associated with the single-linkage clustering algorithm [19]. Therefore, we present both the MST and the hierarchical tree using the single-linkage clustering algorithm. The hierarchical organization found through the MST associated with the distance matrix \(D\) of the distances \(d_{ij}\), shows that is possible to isolate groups which makes sense from an economic point of view by starting from the information carried by the time series.

3.1. Weighted network measures

We also estimate a variety of network measures to characterize the stock market network.\(^3\) In order to study the evolution of the network parameters over time we estimate them using a moving window of fixed size (60 observations, approximately one year of data). Therefore, we estimate the parameters for the sample that comprises observation 1 to 60, 2 to 61, and so forth until we use the entire sample. This dynamic approach allows studying the evolution of the network over time.

---

\(^1\) This dataset was chosen due to the availability of data.

\(^2\) A value of \(\rho_{ij} = 1\) implies that stocks have a perfect correlation between them, whereas a value of \(\rho_{ij} = -1\) suggests that these stocks are perfectly anti-correlated.

\(^3\) See Ref. [22] for a discussion on network measures.
One of the important network measures that we studied is defined as disparity [22]. For a given vertex \( i \) with connectivity \( k_i \) and strength \( s_i \) all weights \( w_{i,j} \) can be of the same order if \( s_i = k_i \). However, if one weight dominates (or a small number of vertices) over the others we may have an heterogeneous case. The disparity of vertex \( i \) can be measured as:

\[
D(i) = \sum_{j \in v(i)} \frac{w_{i,j}^2}{s_i},
\]

where \( v(i) \) is the set of neighbors of \( i \).

Another important network measure is defined as graphcentrality [23]. Graph centrality measures the differences between the centrality of the most central point and that of all others. The larger the relative differences are, the more central a graph (network) is. The general formula for graph centrality is:

\[
C(k) = \frac{1}{\max_{h \in G} d_G(k, h)}.\]

We also study closeness centrality [24], which proxies for the proximity to the rest of vertices in the network. The higher its value, the closer that vertex is to the others (on average). Given a vertex \( k \) and a graph \( G \), it can be defined as:

\[
C(k) = \frac{1}{\sum_{h \in G} d_G(k, h)},
\]

where \( d_G(k, h) \) is the minimum distance from vertex \( k \) to vertex \( h \). This measures the influence of a vertex in a graph.

Domination power measures of individual nodes are able to find the centrality in a network that takes the direction and the weight of the relations into account. Some authors [25] have developed the degree based domination measure called \( \beta \)-measure as described below:

\[
\beta(i) = \sum_{j=1}^{n} \frac{w(i, j)}{\lambda(j)},
\]

where \( w(i, j) \) is the weight of links connected to vertex \( i \) and \( \lambda(j) \) is the dominance weight of node \( j \) given by

\[
\lambda(j) = \sum_{i=1}^{n} w(i, j).
\]

The weighted clustering coefficient [26] can be measured as:

\[
c^{w}_i = \frac{1}{s_i(k_i - 1)} \sum_{j,h} \frac{(w_{i,j} + w_{i,h})}{2} a_{ij} a_{ih} a_{jh},
\]

where \( s_i \) is the vertex strength, \( k_i \) is the degree of vertex \( i \) and \( a_{ij} = 1 \) if there is an edge between \( i \) and \( j \). This coefficient is within the \((0, 1)\) interval. This measures considers not just the number of closed triplets in the neighborhood of a node but also their total relative weight with respect to the strength of the node. The strength measure [22] \( s_i \) is a very significative measure of the network properties in terms of the actual weights and is obtained as:

\[
s_i = \sum_{j \in \gamma(i)} w_{ij},
\]

where the sum runs over the set \( \gamma(i) \) of neighbors of \( i \). The strength of a node integrates the information both with its connectivity and the importance of the weights of its links, and can be considered as the natural generalization of the connectivity.

4. Data

We evaluate 426 observations of 47 stocks using weekly closing prices from January 7, 2000 to February 29, 2008. We build the network using correlation distances between the stock prices in the sample following [1].

We study the role of different stocks in the network built using the previous MST, based on the concept of ultrametricity. We compare the role of stocks by sector and name of stocks. The variable sector, which is categorical, assumes the following categories: Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy. The classification of stocks by their names is provided by the stock exchange.

We also study the dynamics of the evolution of network measures over time. We employ a recursive approach in which we use 60 observations to estimate closeness centrality, disparity, clustering and dominance. In order to analyze the data we average these measures by sector: Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy to compare the sectors relative importance within the network. We calculate these measures using the first 60 observations, the next 60 observations and so forth until we have exhausted the sample.
5. Empirical results

The ultrametric hierarchical tree with the MST makes possible the investigation of the number and nature of the common factors that affect the groups of stocks.

For our full sample we identify the values of $d_{ij}$ between 0 and 1.38 indicating that the Brazilian stock market forms strong clusters and is homogeneous for all characteristics. The division of the full sample in two parts of equal size, does not affect the analysis of the network. The division give us the values of $d_{ij}$ varying from 0 to 1.45, for the first set, and $d_{ij}$ varying from 0 to 1.36, for the second set of stocks. We observe that the values does not change significantly, indicating that the patterns and characteristics of the full sample remains in the divided sample.\(^4\)

The patterns that we can observe are that, if the distances of $d_{ij}$ are low, the groups of stocks tend to form strong clusters and for high values of $d_{ij}$ these links tend to be weak. We can observe that the groups of stocks are homogeneous with respect to sector. It is possible to see that the stocks that belong to the same sector tend to form strong clusters. For example, the stocks of the sector of Materials or the sector of Telecommunications, are highly clustered indicating that these stocks respond to the same economic factors.

In Fig. 1 we present the plot of the MST of a network connecting 47 stocks for the period from January 7, 2000 to February 29, 2008. The stocks are labeled by their industry sectors. The node shape is based on their sectors: (White Ellipse) — Materials; (Black Ellipse) — Telecommunications; (White Triangle) — Industrials; (Black Triangle) — Consumer Discretionary; (White Diamond) — Utilities; (White Square) — Financials; (Black Square) — Energy; (Black Diamond) — Consumer Staples; and (b) Plot of the Taxonomy Hierarchical tree of the subdominant ultrametric associated to the MST of 47 stocks. We present this network based on the names of the sector to which each stock belongs. It presents clusters and clearly shows that stocks of the same sector are clustered together. It also shows that the clusters are homogeneous with respect to the sector. (See Fig. 2.)

Fig. 3 presents the plot of the average dominance for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. It is worth noticing that different sectors have

\(^4\) We also split the sample into two (first 213 and last 213 observations) and plot both the MST and the taxonomy hierarchical tree of the subdominant ultrametric associated to the MST and find qualitatively the same results. These results are not shown to conserve space.
very different dominance. The difference between the sector with the highest average dominance (Materials) and the lowest average dominance (Industrial) is large. It is more than four times larger, which suggests that the relationship of different stocks within the network is highly heterogeneous. Fig. 4 presents the average clustering coefficients for all sectors. These figures were done using the full sample. In this case the ranking changes if compared to Fig. 3 and the heterogeneity is substantially reduced.

In Fig. 5 we present the average closeness centrality for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. The Financial sector has the largest average closeness centrality. This could be interpreted as an important role for the financial sectors within the network. Most of the stocks that belong to this sector are banks. Bank lending is one of the most important sources of funding for Brazilian firms, which suggests that the banking sector should be strongly connected to most of the other sectors. Fig. 6 presents the average disparity for the same sectors. This measure suggests that the energy sector is the most heterogeneous within the sample. In this case, firms belong to the oil and electricity sub-sectors, which may explain to some degree the larger heterogeneity.

Fig. 7 presents the plot of the time varying closeness centrality for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. In the beginning of the sample the closeness centrality fluctuates around 0.2 but it has a peak in the 2004 period, which is followed by a more volatile period with further peaks in 2006 and 2008. Fig. 8 presents the plot of the time varying graph centrality for all sectors. Again, we observe a peak in the 2004 period. In order to calculate the time varying measures we employ 60 weekly observations (approximately one year of data) for each calculation and estimate these measures recursively on a moving average basis. Therefore, as of March 2004 the measures were calculated using data from March 2003 to March 2004 (approximately).
One of the reasons that may explain these peaks is the change in the elections that occurred in 2002, which were followed by a period of high exchange rate volatility, large declines in stock markets in the end of 2002 and beginning of 2003 and a subsequent phase of euphoria.

Fig. 9 presents the plot of the time varying disparity measures for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. In this case the peaks in the 2004 period are not present. Fig. 10 presents the time varying clustering coefficients for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. There is a downward trend for all sectors, within the sample period.

Fig. 11 presents the time varying average value of strength measure for all sectors. In this case we see major changes in 2004 and subsequent peaks in 2005, 2006 and 2007. The differences between sectors are large, which suggests that some sectors are more interconnected within the network than others. This is an important empirical results for the development of portfolio strategies. The larger the strength of a specific sector the larger should be its exposure to shocks in the stock market. Furthermore, if specific sectors have a larger impact in specific periods it is crucial to take into account these possible changes when defining investment strategies.

In Fig. 12 we present the plot of the time varying dominance measure for the Materials, Consumer Staples, Financials, Telecommunication Services, Utilities, Industrials, Consumer Discretionary and Energy sectors. In this case, it is possible to see a different dynamics that governs these sectors.
Fig. 7. Time varying closeness centrality.

Fig. 8. Time varying graph centrality.

Table 1
Brazilian stock sectors.

<table>
<thead>
<tr>
<th>Sector</th>
<th>Closeness centrality</th>
<th>Graph centrality</th>
<th>Dominance</th>
<th>Clustering</th>
<th>Disparity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials</td>
<td>0.0339</td>
<td>0.5844</td>
<td>2.1094</td>
<td>0.0605</td>
<td>0.0352</td>
</tr>
<tr>
<td>Cons. staples</td>
<td>0.0330</td>
<td>0.6052</td>
<td>0.9693</td>
<td>0.0580</td>
<td>0.0260</td>
</tr>
<tr>
<td>Financials</td>
<td>0.0384</td>
<td>0.5856</td>
<td>1.2084</td>
<td>0.0695</td>
<td>0.0306</td>
</tr>
<tr>
<td>Telecomm.</td>
<td>0.0229</td>
<td>0.6739</td>
<td>1.1139</td>
<td>0.0420</td>
<td>0.0235</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.0322</td>
<td>0.6323</td>
<td>1.3407</td>
<td>0.0541</td>
<td>0.0252</td>
</tr>
<tr>
<td>Industrials</td>
<td>0.0309</td>
<td>0.5888</td>
<td>0.3249</td>
<td>0.0523</td>
<td>0.0245</td>
</tr>
<tr>
<td>Cons. discretionary</td>
<td>0.0290</td>
<td>0.5747</td>
<td>0.5231</td>
<td>0.0519</td>
<td>0.0272</td>
</tr>
<tr>
<td>Energy</td>
<td>0.0357</td>
<td>0.5898</td>
<td>0.4103</td>
<td>0.0644</td>
<td>0.0509</td>
</tr>
</tbody>
</table>

This table shows each sector that comprises our sample and the average values of each network parameter.

In Table 1 we can observe the stock sectors and the average values of measures of network. It is evident that different sectors have diverse roles within the network. Overall, the empirical findings suggests that the role of specific sectors change over time. Also that it is crucial to follow up the dynamics of the network topology to understand which are the sector with higher exposure to shocks.
6. Conclusions

This paper shows that the MST and the ultrametric hierarchical tree can be used to analyze the stock prices of the Brazilian financial sector. The evidence suggests that the Brazilian equity sector forms clusters based on the industry and is homogeneous. We split the sample into two different periods and the qualitative results obtained from the MST and the hierarchical tree are robust.
Furthermore, network measures suggest that the Financial, Material and Energy sectors are relatively more important within the network with a high influence on other sectors. Although the telecommunications sector has been very active in the recent years, with introduction of new technologies and so forth it does not have a very important role in the network. Therefore, trading strategies should place a larger weight on these sectors. These findings imply that specific sectors may play a dominant role within stock market networks. Further research could exploit this issue by making international comparisons.

An important contribution is that we show that some of the network characteristics seem to be changing over time. Furthermore, the relative importance of specific sectors also seems to be changing. This suggests that these tools should be
used on a continued basis to help the portfolio and risk management processes. Further research could exploit the dynamic changes that have occurred in the recent crisis and how the role of the banking sector has evolved.

Acknowledgements

The authors thank the suggestions provided by the referees, which helped to improve the paper. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the Banco Central do Brasil. Benjamin M. Tabak and Daniel O. Cajueiro gratefully acknowledge financial support from CNPQ Foundation.

References