To lag or not to lag?

Leonidas Sandoval Junior

Insper, Instituto de Ensino e Pesquisa

May 16, 2012

Abstract

Financial markets worldwide do not have the same working hours. As a consequence, the study of correlation or causality between financial market indices becomes dependent on wether we should consider in computations of correlation matrices all indices in the same day or lagged indices. The answer is that we should consider both.

I. Introduction

Probably the major issue when dealing with stock markets around the globe is how to treat the differences in operating times of the stock exchanges. It is quite clear that financial markets in Pacific Asia often influence financial markets in the American continent, and that the New York Stock Exchange influences the next day of the Tokyo Stock Exchange, but how can one deal with these influences when, for example, one is trying to build a portfolio of international assets by minimizing the covariance matrix built from their time series? Which time series should be lagged with respect to others, if any?

Section II explains which data is being used and some of the methodology. Section III performs a comparison between the differences of opening hours of stock exchanges worldwide the correlations of the log-returns of their indices. Section IV analyzes the structure of the probability distribution of the eigenvalues of the eigenvalues obtained from the correlation matrix of the log-returns extracted from time series that span the years from 2007 to 2011. Section V employs an enlarged correlation matrix consisting of lagged and non-lagged indices and builds a map of the resulting network. The conclusions are drawn in section VI.

II. Data and methodology

We consider the time series of 92 benchmark indices of stock exchanges worldwide. The data span the period that goes from 2007 to 2011, and the indices analyzed are the S&P 500 and Nasdaq, both of the USA, the indices from Canada and Mexico (North America), Panama, Costa Rica, Bermuda, and Jamaica (Central America and the Caribbean), Brazil, Argentina, Chile, Colombia, Venezuela, and Peru (South America), the UK, Ireland, France, Germany, Switzerland, Austria, Italy, Malta, Belgium, the Netherlands, Luxembourg, Sweden, Denmark, Finland, Norway, Iceland, Spain, Portugal, Greece, the Czech Republic, Slovakia, Hungary, Serbia, Croatia, Slovenia, Bosnia & Herzegovina, Montenegro, Macedonia, Poland, Romania, Bulgaria, Estonia, Latvia, Lithuania, and Ukraine (Europe), Russia, Kazakhstan, and Turkey (Eurasia), Cyprus, Israel, Palestina, Lebanon, Jordan, Saudi Arabia, Kuwait, Bahrain, Qatar, the United Arab Emirates, Ohman, Pakistan, India, Sri Lanka, Bangladesh, Japan, Hong Kong, China, Mongolia, Taiwan, South Korea, Thailand, Viet Nam, Malaysia, Singapore, Indonesia, and the Philipines (Asia), Australia and New Zealand (Oceaina), Moroco, Tunisia, Egypt, Ghana, Nigeria, Kenya, Tanzania, Namibia, Botswana, South Africa, and Mauritius islands (Africa).

The log-returns of these indices, defined as

$$r_t = \ln(P_t) - \ln(P_{t-1}) \ , \tag{1}$$

where P_t is the closing price of the index at day t and P_{t-1} is the closing price of the same index at day $t-1$, are used in order to compute correlation matrices between all log returns. There are many correlation measurements available, being the Pearson correlation the most used one, but our choice was the Spearman rank correlation, since it is better suited to analize nonlinear correlations.

One may then calculate the eigenvalues and eigenvalues of the resulting correlation matrices, and comparing them with the results obtained by a correlation matrix resulting from shuffled data, in order to pinpoint differences between them. The correlation matrix may also be used in order to compute a distance measure between the indices, and this distance matrix may then be used to draw a map of the network of indices.

III. Correlations and working hours

Figure 1 shows the opening and closing hours of the 92 stock exchanges considered in this letter, not taking into account possible differences in intervals of the year due to daylight saving time. The dashed line marks the opening and closing hours of the New York Stock Exchange (NYSE), which shall be considered as our benchmark. Considering data from the first semester of 2008, as an example, we may calculate the correlation matrices of all indices with the S&P 500 on the same day and of all indices with the S&P 500, but now all lagged by one day. We use Spearman rank correlation, for it captures best nonlinear relationships, but it can be shown empirically [1] that the Pearson correlation is as good as the Spearman one in what concerns our set of data.

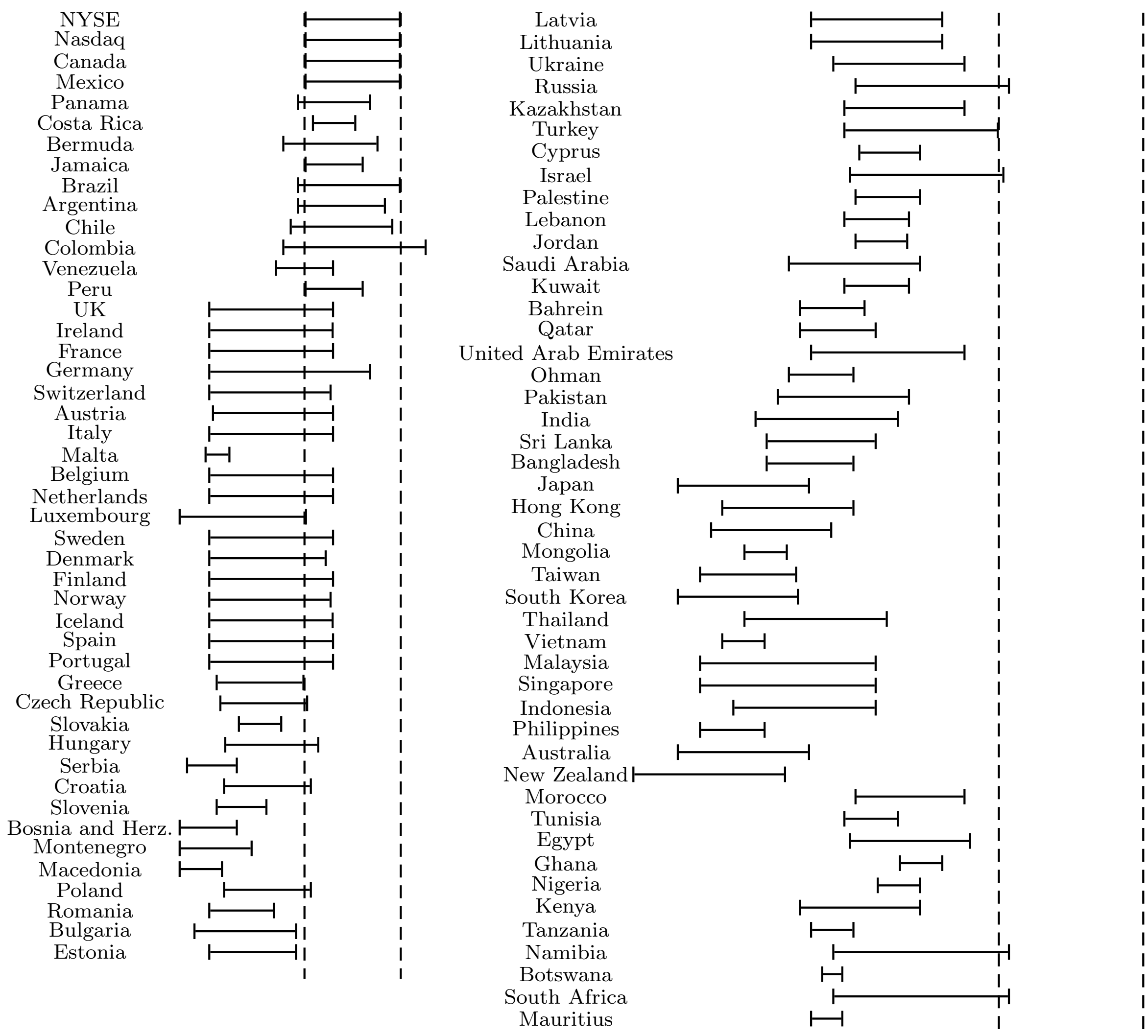

Fig. 1. Opening and closing hours of some of the main stock exchanges in the world. The dashed line marks the opening and closing hours of the New York Stock Exchange (NYSE).

By doing so, we may notice that some correlations between the S&P 500 and the other indices actually increase when we lag those indices by one day. In contrast, the correlation with other indices drop when we consider lagged data. A brief comparison of the difference between lagged and non lagged correlations reveal that most indices which have positive results for lagged correlation are those of countries whose intersection with the opening hours of the NYSE is inexistent. For most of the other indices whose stock exchanges operate partially or totally in the same hours of the NYSE, correlation drops when one considers lagged data. One may even set the rule that, when one market's working hours do not overlap with the ones of the NYSE, then lagged correlation is best than the non lagged one. For the first semester of 2008, about 86% of market indices satisfy this rule. The exceptions are those from Panama, Bermuda, Colombia, Venezuela, Peru, Luxembourg, Norway, Iceland, Czech Republic, Croatia, Poland, Russia, Namibia, and South Africa. These markets have either a low volume of trading or very small intersecting hours with the NYSE.

The same percentage holds if one considers data for the second semester of 2008, which presented much higher volatility than the first semester of the same year: 86% of the indices conform to the rule that lagged correlations are best for the stock exchanges that operate at different hours from the NYSE. The exceptions are now the ones from Panama, Bermuda, Jamaica, Venezuela, Luxembourg, Denmark, Iceland, Portugal, Czech Republic, Hungary, Poland, Russia, Namibia, and South Africa. Table 1 shows the percentage of nodes that conform to the rule, and the exceptions to it, for intervals of six months data concerning the years from 2007 to 2011.

Semester/Year	Conformity	Exceptions
01/2007	90%	Panama, Bermuda, Venezuela, Ireland, Austria, and Luxembourg, Iceland, Czech Republic, Hungary, and Croatia
02/2007	94%	Panama, Bermuda, Jamaica, Venezuela, Iceland, and Croatia
01/2008	86%	Panama, Bermuda, Colombia, Venezuela, Peru, and Luxembourg, Norway, Iceland, Czech Republic, Croatia, Poland, and Russia, Namibia and South Africa
02/2008	86%	Panama, Bermuda, Jamaica, Venezuela, Luxembourg, and Denmark, Iceland, Portugal, Czech Republic, Hungary, Poland, and Russia, Namibia and South Africa
01/2009	93%	Bermuda, Brazil, Peru, Hungary, Croatia, Russia, and South Africa
02/2009	94%	Belgium, Spain, Hungary, Croatia, Poland, and Namibia
01/2010	91%	Costa Rica, Jamaica, Peru, Belgium, Spain, and Hungary, Croatia, Poland, and Russia
02/2010	90%	Costa Rica, Bermuda, Jamaica, Brazil, Spain, and Hungary, Croatia, Poland, Namibia, and South Africa
01/2011	99%	Panama and Venezuela
02/2011	98%	Panama, Costa Rica, and Jamaica

Table 1. Percentage of nodes that conform to the rule that lagged correlations of stock martket indices with the S&P 500 are best when there is no intersection between opening hours, and the exceptions to it, for intervals of six months.

Most of the indices conform to the rule and, with the exception of those indices from the Caribbean, most of those have very small intersections with the NYSE. Notable exceptions are Brazil and Spain, from 2009 to 2010, both presenting small correlations with the NYSE for that period of time, probably due to particularities in their stock exchanges at the time. In the case of Brazil, the stock market has been doing better than the NYSE in the past few years, due to a great increase in foreign investment, since the country is almost untouched by the international economic crisis following 2008.

IV. Eigenvalue probability distribution

Another way to analyze the influence of different opening hours in the correlation between the time series of international stock exchange indices is by considering the eigenvalues of the correlation matrix. Borrowing results from [2] and [3], if we calculate the correlation concerning the 92 indices here considered using data for the period 2007 to 2011, we then obtain a frequency distribution of the eigenvalues of such a matrix. This distribution is quite different from one that may be obtained by randomly shufling the original time series so as to preserve their averages and standard deviations, but destroy any simultaneous relation between each time series. The random distribution approaches the theoretical result obtained using Random Matrix Theory [4]. Figure 2 shows the eigenvalues of the correlation matrix for real data (solid bars) and the region where eigenvalues from 1000 simulations with randomized data may be found (gray area).

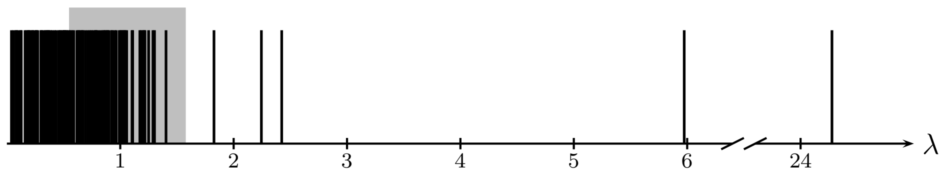

Fig. 2. Eigenvalues of the correlation for data from 2007 to 2011, in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

Note that there are two eigenvalues that stand out from the others. The highest eigenvalue is associate with the "market mode", which is a common movement of all the stocks. This market mode has been observed in a variety of stock markets and economic ones, and also in the study of stock market indices ([5] and references therein). Now, the second eigenvalue has a characteristic that is unique to financial market indices of stock exchanges that operate at different times. Analyzing the eigenvector associated with the first and the second highest eigenvalues, Figure 3, one may notice that it the eigenvectors for the highest eigenvalue are almost all positive, and most give the same amount of contribution; now the eigenvector for the second highest eigenvalue separates two types of sets of indices. In the figure, white bars correspond to positive signs in the eigenvector, and gray bars correspond to negative signs in the eigenvector.

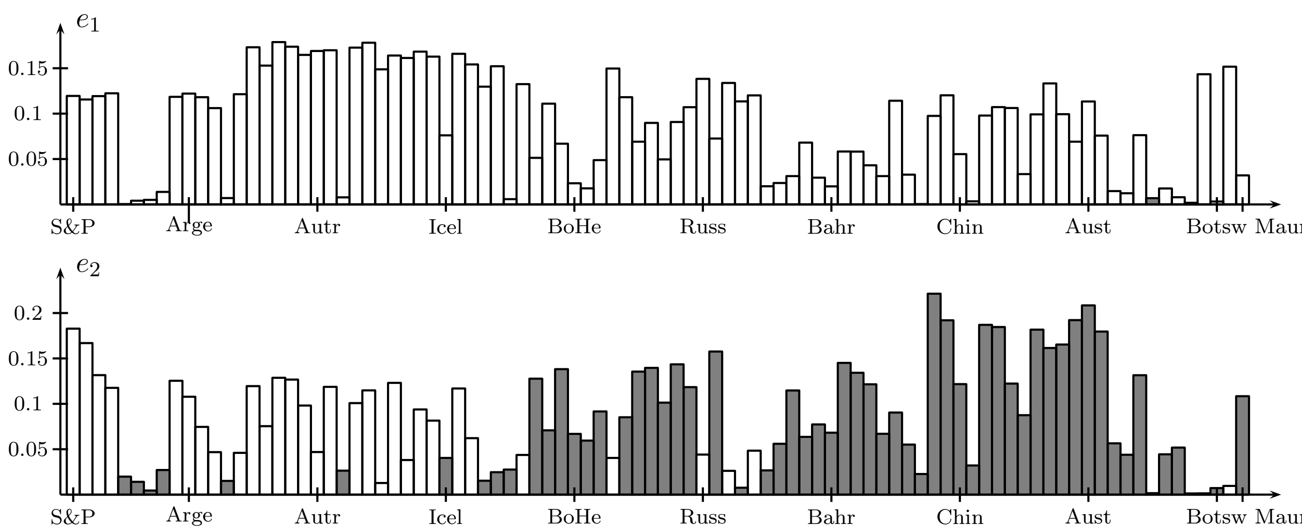

Fig. 3. Contributions of the stock market indices to eigenvectors e_1 and e_2, corresponding, respectively, to the first and the second largest eigenvalues of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to the period 2007-2010. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, Pana, CoRi, Berm, Jama, Braz, **Arge**, Chil, Colo, Vene, Peru, UK, Irel, Fran, Germ, Swit, **Autr**, Ital, Malt, Belg, Neth, Luxe, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, Slok, Hung, Serb, Croa, Slov, **BoHe**, Mont, Mace, Pola, Roma, Bulg, Esto, Latv, Lith, Ukra, **Russ**, Kaza, Turk, Cypr, Isra, Pale, Leba, Jord, SaAr, Kuwa, **Bahr**, Qata, UAE, Ohma, Paki, Indi, SrLa, Bang, Japa, HoKo, **Chin**, Mong, Taiw, SoKo, Thai, Viet, Mala, Sing, Indo, Phil, **Aust**, NeZe, Moro, Tuni, Egyp, Ghan, Nige, Keny, Tanz, Nami, **Bots**, SoAf, and **Maur**.

One can see that most indices that go from Greece eastwards, in terms of longitude, present opposite values for the eigenvalue from the one of the NYSE. Exceptions are Panama, Costa Rica, Bermuda, Jamaica, Venezuela, Malta, Iceland, Russia, Turkey, Israel, and South Africa. All of these indices appear with a very small value in Figure 3, and most of them figure in Table 1 as well. Many of them are small markets or have small intersections with the NYSE (timewise). If one laggs those indices that appear with negative values in the eigenvector corresponding to the second highest eigenvalue of the correlation matrix, one then obtains a new structure where the highest eigenvalue still corresponds to a market mode and the second highest eigenvalue is now associated with a new structure, where European indices detach from the others, indicating an internal structure of the world financial markets.

V. Large correlation matrix

So, lagging some indices may give better results when analyzing the structure of stock market relations, but it is not optimal. A new alternative is proposed in this letter, which is to consider the time series of the international market indices and the time series of the same indices, but lagged by one day. So, instead of 92 indices, we now have 184 of them, counting the lagged and non lagged ones. Building a correlation matrix based on the 184 resulting time series, one obtains very interesting results. First, we plot in Figure 4 the eigenvalues associated with the correlation matrix for the whole data (2007 to 2010), together with the shaded region corresponding to the eigenvalues of 100 simulations with randomized data.

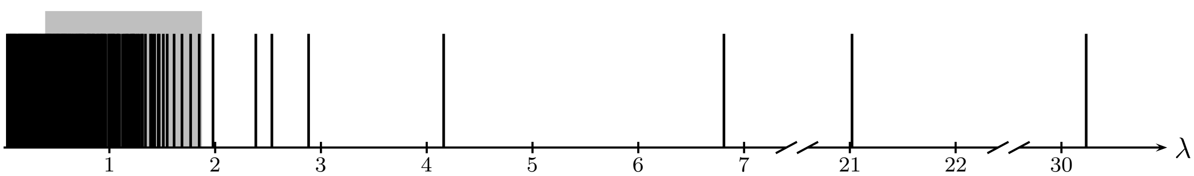

Fig. 4. Eigenvalues of the correlation for data from 2007 to 2011 with lagged values, in order of magnitude. The shaded area corresponds to the eigenvalues predicted for randomized data.

The highest eigenvalue again corresponds to the market mode, but the second highest one reveals a different structure. Figure 5 shows the elements of the eigenvector associated with this second highest eigenvalue, with gray bars meaning negative results and white bars meaning positive values. There is a clear division between non lagged (most of the gray bars) and lagged (most of the white bars) indices, with some exceptions, which are, most of the time, the same as the ones in Table 1.

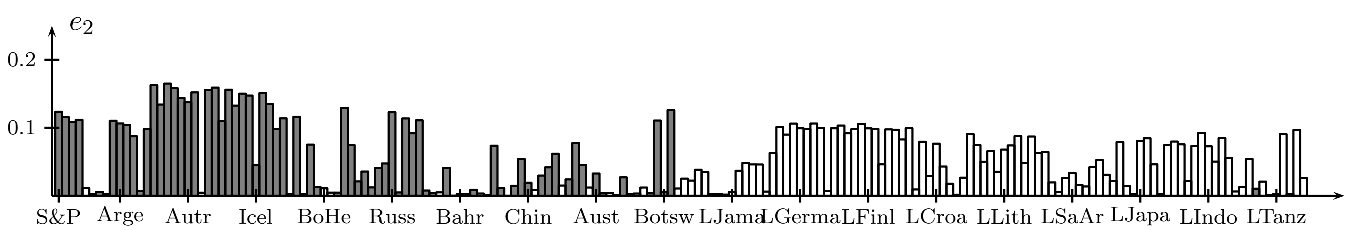

Fig. 5. Contributions of the stock market indices to eigenvector e_2, corresponding to the second largest eigenvalue of the correlation matrix. White bars indicate positive values, and gray bars indicate negative values, corresponding to 2007-2010. The indices are aligned in the following way: **S&P**, Nasd, Cana, Mexi, Pana, CoRi, Berm, Jama, Braz, **Arge**, Chil, Colo, Vene, Peru, UK, Irel, Fran, Germ, Swit, **Autr**, Ital, Malt, Belg, Neth, Luxe, Swed, Denm, Finl, Norw, **Icel**, Spai, Port, Gree, CzRe, Slok, Hung, Serb, Croa, Slov, **BoHe**, Mont, Mace, Pola, Roma, Bulg, Esto, Latv, Lith, Ukra, **Russ**, Kaza, Turk, Cypr, Isra, Pale, Leba, Jord, SaAr, Kuwa, **Bahr**, Qata, UAE, Ohma, Paki, Indi, SrLa, Bang, Japa, HoKo, **Chin**, Mong, Taiw, SoKo, Thai, Viet, Mala, Sing, Indo, Phil, **Aust**, NeZe, Moro, Tuni, Egyp, Ghan, Nige, Keny, Tanz, Nami, **Bots**, SoAf, Maur, and their lagged versions.

The true new information, though, appears when we build asset graphs [6]-[14] based on a distance measure built from the correlation between indices. Let us explain some details, first. We build a distance matrix from the correlation matrix of a sequence of time series using the following definition of distance:

$$d_{ij} = 1 - c_{ij} \ , \tag{2}$$

where d_{ij} is the distance between nodes i and j and c_{ij} is the correlation between the same two indices. As correlations vary from -1 (anticorrelated) to 1 (completely correlated), the distances vary from 0 (totally correlated) to 2 (completely anticorrelated). Totally uncorrelated indices would have distance 1 between them.

Now, this distance matrix may be used in order to establish a threshold value T for the distance. An asset graph will be defined as the network obtained by eliminating all connections corresponding to distances bellow T and then eliminating all nodes which are not connected. A representation of N nodes in a Euclidena space would need to have N dimensions in order to be a perfect one, but usually representations in lower dimensions also offer a good agreement. By using a process called Classical Multidimensional Scaling [15], we may assign coordinates in a three dimensional space such that the distances between nodes are a good approximation of the real distances between them.

Figure 6 shows the asset graph thus built from data collected from 2007 to 2011 and distance threshold $T = 0.6$. One can see the existence of two clusters, separated by few connections. The connections that link both clusters are the ones from Pacific Asia and Oceania, and the figure displays the indices from East to West, almost like in a world map. The connection between non lagged and lagged clusters is lost at $T = 0.5$. A similar graph may be obtained if we also consider indices lagged by two days. The result are three clusters, connected by the Pacific Asian and Oceanian indices. For indices lagged by three days, we have four clusters.

This feature is peculiar to data related with world stock market indices. Doing the same analysis for data related with a stock exchange, a single cluster with mixed lagged and non lagged indices is obtained. It clearly indicates a chain of correlations, and we can hypothesize a causality structure going from East to West, and then back to the East, in a cyclic pattern.

The same type of asset graphs can be built for other periods of time, without much change. The network structure persists and is robust in time, and different levels of threshold may be used in order to analyze clusters formed at different strenghts of connections [2],[3].

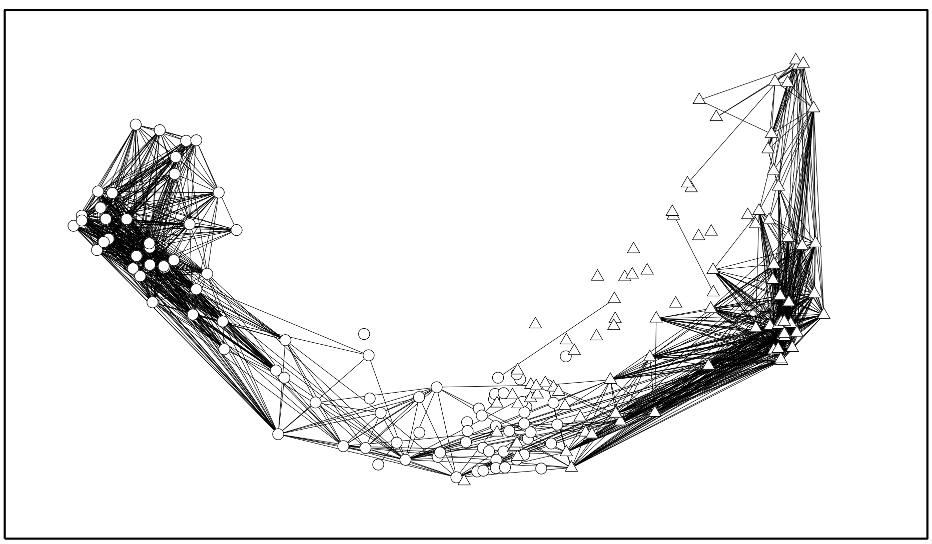

Fig.6. Asset tree for data from 2007 to 2011 for $T = 0.6$. Non-lagged indices appear as circular dots, and lagged indices appear as triangles.

VI. Conclusions

What we may then conclude from this discussuion is that, if one wants to study the relations and dynamics of the world stock market indices, one must then consider the cyclic chain of influences, and probably decide to adopt a model where there are both non lagged and lagged indices.

The author acknowledges the support of this work by a grant from Insper, Instituto de Ensino e Pesquisa. This letter was written using LaTeX, all figures were made using PSTricks, and the calculations were made using Matlab and Excel. All data are freely available upon request on leonidassj@insper.edu.br.

References

[1] L. Sandoval Jr., *Pruning a Minimum Spanning Tree*, Physica A 391 (2012) 2678–2711.

[2] L. Sandoval Jr., *Cluster formation and evolution in networks of financial market indices*, (2011) arXiv:1111.5069v1.

[3] L. Sandoval Jr., *Survivability and centrality measures for networks of financial market indices*, (2012) arXiv:1201.4490v1.

[4] M. L. Mehta, *Random Matrices*, (2004) Academic Press.

[5] L. Sandoval Jr. and I. De P. Franca, *Correlation of financial markets in times of crisis*, Physica A 391, (2012) 187-208.

[6] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertész, *Dynamic asset trees and portfolio analysis*, The European Phys. J. B 30 (2002) 285-288.

[7] J.-P. Onnela, A. Chakraborti, and K. Kaski, *Dynamics of market correlations: taxonomy and portfolio analysis*, Phys. Rev. E 68 (2003) 1-12.

[8] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertész, *Dynamic asset trees and Black Monday*, Physica A 324 (2003) 247-252.

[9] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Asset trees and asset graphs in financial markets*, Phys. Scripta T 106 (2003) 48-54.

[10] J.-P. Onnela, K. Kaski, and J. Kertész, *Clustering and information in correlation based financial networks*, Eur. Phys. J. B 38 (2004) 353-362.

[11] S. Sinha and R.K. Pan, *Uncovering the internal structure of the Indian financial market: cross-correlation behavior in the NSE*, in "Econophysics of markets and business networks", Springer (2007) 215-226.

[12] M. Ausloos and R. Lambiotte, *Clusters or networks of economies? A macroeconomy study through gross domestic product*, Physica A 382 (2007) 16-21.

[13] L. Sandoval Jr., *A Map of the Brazilian Stock Market*, Advances in Complex Systems 15 (2012) 1250042.

[14] T. Heimo, K. Kaski, and J. Saramäki, *Maximal spanning trees, asset graphs and random matrix denoising in the analysis of dynamics of financial networks*, Physica A 388 (2009) 145-156.

[15] I. Borg and P. Groenen, *Modern Multidimensional Scaling: theory and applications*, 2nd edition, (2005) Springer-Verlag.