# ISCIES'11

## Proceedings of ISCIES'11
## 2nd International Symposium on Computational Intelligence for Engineering Systems



**Edição Científica:** Ana Madureira (ISEP), Cecília Reis (ISEP), Viriato Marques (ISEC)

**Editor:** ISEC (Instituto Superior de Engenharia de Coimbra)

**ISBN: 978-989-8331-12-0**

isec
Instituto Superior de Engenharia de Coimbra

Coimbra, Portugal
16-18 November, 2011
http://www.iscies11.ipc.pt

isep Instituto Superior de Engenharia do Porto

gecad
Grupo de Investigação em Engenharia do Conhecimento e Apoio à Decisão

ISEL
INSTITUTO SUPERIOR DE ENGENHARIA DE LISBOA

IEEE
PORTUGAL SECTION

IEEE Computational Intelligence Society
Portugal Chapter

# Visualization of Relations Between Financial Indices Using Multidimensional Scalling

J. Tenreiro Machado
Dept. of Electrical Engineering
Institute of Engineering
Porto, Portugal
Tel.: +(351)228340500
Fax: +(351)228321159
Email: jtm@isep.ipp.pt

Gonçalo M. Duarte
Lusofona University
Lisboa, Portugal
Tel.: +(351)216001515
Email: monteiro.duarte@gmail.com

Fernando B. Duarte
Dept. of Mathematics
School of Technology
Viseu, Portugal
Tel.: +(351)232480500
Fax: +(351)232424651
Email: fduarte@estv.ipv.pt

***Abstract*—This paper applies Multidimensional scaling techniques for visualizing possible time-varying correlations between twenty five stock market values. The method is useful for observing stable or emerging clusters of stock markets with similar behavior. The graphs may also guide the construction of multivariate econometric models.**

## I. Introduction

It seems that there are many distinct analogies between the dynamics of complex physical and economical or even social systems. The methods and algorithms that have been explored for description of physical phenomena become an effective background and inspiration for very productive methods used in the analysis of economical data [1, 2, 3, 4].

Economical indices measure the performance of segments of the stock market and are normally used to benchmark the performance of stock portfolios. This paper proposes a descriptive method which analyzes possible correlations in international stock markets. The study of the correlation of international stock markets may have different motivations. Economic motivations to identify the main factors which affect the behavior of stock markets across different exchanges and countries. Statistical motivations to visualize correlations in order to suggest some potentially plausible parameter relations and restrictions. The understanding of such correlations would be helpful to the design good portfolios [1, 2].

Bearing these ideas in mind the outline of our paper is as follows. In Section 2 we give the fundamentals of the multidimensional scaling (MDS) technique, which is the core of our method, and we discuss the details that are relevant for our specific application. In Section 3 we apply our method for daily data on twenty five stock markets, including major American, Asian/Pacific, and European stock markets. In Section 4 we conclude the paper with some final remarks and potential topics for further research.

## II. Fundamental concepts

MDS is a set of data analysis techniques for analysis of similarity or dissimilarity data. It is used to represent (dis)similarity data between objects by a variety of distance models.

The term similarity is used to indicate the degree of “likeness” between two objects, while dissimilarity indicates the degree of “unlikeness”. MDS represents a set of objects as points in a multidimensional space in such a way that the points corresponding to similar objects are located close together, while those corresponding to dissimilar objects are located far apart. The researcher then attempts to “make sense” of the derived object configuration by identifying meaningful regions and/or directions in the space.

In this article, we introduce the basic concepts and methods of MDS. We then discuss a variety of (dis)similarity measures and the kinds of techniques to be used. The main objective of MDS is to represent these dissimilarities as distances between points in a low dimensional space such that the distances correspond as closely as possible to the dissimilarities.

Let $n$ be the number of different objects and let the dissimilarity for objects $i$ and $j$ be given by $\delta_{ij}$. The coordinates are gathered in an $n \times p$ matrix $\mathbf{X}$, where $p$ is the dimensionality of the solution to be specified in advance by the user. Therefore, row $i$ from $\mathbf{X}$ gives the coordinates for object $i$. Let $d_{ij}$ be the Euclidean distance between rows $i$ and $j$ of $\mathbf{X}$ defined as

$$d_{ij} = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2} \qquad (1)$$

that is, the length of the shortest line connecting points $i$ and $j$. The objective of MDS is to find a matrix $\mathbf{X}$ such that $d_{ij}$ matches $\delta_{ij}$ as closely as possible. This objective can be formulated in a variety of ways but here we use the definition of raw-Stress $\sigma^2$, that is,

$$\sigma^2 = \sum_{i=2}^{n} \sum_{j=1}^{i-1} w_{ij} \left(\delta_{ij} - d_{ij}\right)^2 \qquad (2)$$

by Kruskal [5] who was the first one to propose a formal measure for doing MDS. This measure is also referred to as the least-squares MDS model. Note that due to the symmetry of the dissimilarities and the distances, the summation only involves the pairs $i, j$ where $i > j$. Here, $w_{ij}$ is a user defined weight that must be nonnegative. The minimization of $\sigma^2$ is a complex problem. Therefore, MDS programs use iterative numerical algorithms to find a matrix $\mathbf{X}$ for which $\sigma^2$ is a minimum. In addition to the raw stress measure there exist other measures for doing stress. One of them is normalized raw stress, which is simply raw stress divided by the sum of squared dissimilarities. The advantage of this measure over raw stress is that its value is independent of the scale and the number of dissimilarities. The second measure is Kruskal's stress-1 which is equal to the square root of raw stress divided by the sum of squared distances. A third measure is Kruskal's stress-2, which is similar to stress-1 except that the denominator is based on the variance of the distances instead of the sum of squares. Another measure that seems reasonably popular is called S-stress and it measures the sum of squared error between squared distances and squared dissimilarities.

Because Euclidean distances do not change under rotation, translation, and reflection, these operations may be freely applied to MDS solution without affecting the raw-stress. Many MDS programs use this indeterminacy to center the coordinates so that they sum to zero dimension wise. The freedom of rotation is often exploited to put the solution in so-called principal axis orientation. That is, the axis are rotated in such a way that the variance of $\mathbf{X}$ is maximal along the first dimension, the second dimension is uncorrelated to the first and has again maximal variance, and so on.

In order to assess the quality of the MDS solution we can study the differences between the MDS solution and the data. One convenient way to do this is by inspecting the so-called Shepard diagram. A Shepard diagram shows both the transformation and the error. Let $p_{ij}$ denote the proximity between objects $i$ and $j$. Then, a Shepard diagram plots simultaneously the pairs $(p_{ij}, d_{ij})$ and $(p_{ij}, \delta_{ij})$. By connecting the $(p_{ij}, \delta_{ij})$ points a line is obtained representing the relationship between the proximities and the disparities. The vertical distances between the $(p_{ij}, \delta_{ij})$ points and $(p_{ij}, d_{ij})$ are equal to $\delta_{ij} - d_{ij}$, that is, they give the errors of representation for each pair of objects. Hence, the Shepard diagram can be used to inspect both the residuals of the MDS solution and the transformation. Outliers can be detected as well as possible systematic deviations.

Measuring and predicting human judgment is an extremely complex and problematic task. There have been many techniques developed to deal with such type of problems. These techniques fall under a generic category called Multidimensional Scaling (MDS). Generally speaking MDS techniques develop spatial representations of psychological stimuli or other complex objects about which people make judgments (e.g. preference, relatedness), that is they represent each object as a point in a *n*-dimensional space. What distinguishes MDS from other similar techniques (e.g. factor analysis) is that in MDS there are no preconceptions about which factors might drive each dimension. Therefore, the only data needed is a measure for the similarity between each possible pair of objects under study. The result is the transformation of the data into similarity measures which can be represented by Euclidean distances in a space of unknown dimensions [6]. The greater the similarity of two objects, the closer they are in the *n*-dimensional space. After having the distances between all the objects, the MDS techniques analyze how well they can be fitted by spaces of different dimensions. The analysis is normally made by gradually increasing the number of dimensions until the quality of fit (measured for example by the correlation between the data and the distance) is little improved with the addition of a new dimension. In practice a good result is normally reached well before the number of dimensions theoretically needed to a perfectly fit is reached (i.e. $N-1$ dimensions for $N$ objects) [7, 8, 9, 10].

In the MDS method a small distance between two points corresponds to a high correlation between two stock markets and a large distance corresponds to low or even negative correlation [11, 12]. A correlation of one should lead to zero distance between the points representing perfectly correlated stock markets. MDS tries to estimate the distances for all pairs of stock markets to match the correlations as close as possible. MDS may thus be seen as an exploratory technique without any distributional assumptions on the data. The distances between the points in the MDS maps are generally not difficult to interpret and thus may be used to formulate more specific models or hypotheses. Also, the distance between two points should be interpreted as being the distance conditional on all the other distances. One possibility to obtain such an approximate solution is given by minimizing the stress function. The obtained representation of points is not unique in the sense that any rotation or translation of the points retains the distances [13].

## III. Analysis of financial indices

In this section we study numerically the twenty five selected stock markets, including seven American, eleven European and seven Asian/Pacific markets.

Our data consist of the $n$ daily close values of $S = 25$ stock markets,listed in Table I, from January 2, 2000, up to December 31, 2009, to be denoted as $x_i(t), t = 1, \cdots n$, $i = 1, \cdots, S$.

The data are obtained from data provided by Yahoo Finance web site [14] and [15] , and they measure indices in local currencies.

Figure 1 depicts the time evolution, of daily, closing price of the twenty five stock markets versus year with the well-know noisy and "chaotic-like" characteristics.

The section is organized in four subsections, each adopting as "similarity measure" one correlation's coefficient. The first adopts an analysis based on a squared cosine coefficient correlation ($\xi$), the second adopts an analysis based on the Pearson coefficient correlation ($\rho$), the third adopts the Kendall-tau coefficient of correlation ($\tau$) and the four adopts the Spearman correlation's coefficient ($\sigma$).

### A. MDS analysis based on squared cosine correlation

In this subsection we apply the MDS method using as "similarity measure" the values of squared cosine correlation, $\xi(i, j)$, of all the stock markets with the daily close values.

Table I
TWENTY FIVE STOCK MARKETS

| $i$ | Stock market index | Abbrev. | Country |
|---|---|---|---|
| 1 | Dutch Euronext Amsterdam | aex | Netherlands |
| 2 | Index of the Vienna Bourse | atx | Austria |
| 3 | EURONEXT BEL-20 | bfx | Belgium |
| 4 | Bombay Stock Exchange Index | bse | India |
| 5 | SÃ£o Paulo (Brazil) Stock | bvsp | Brazil |
| 6 | Budapest Stock Exchange | bux | Hungary |
| 7 | Dow Jones Industrial | dji | USA |
| 8 | Cotation AssistÃ©e en Continu | cac | France |
| 9 | Footsie | ftse | United Kingdom |
| 10 | Deutscher Aktien Index | dax | Germany |
| 11 | Standard & Poor's | sp500 | USA |
| 12 | Toronto Stock Exchange | tsx | Canada |
| 13 | Stock market index in Hong Kong | hsi | Hong Kong |
| 14 | Iberia Index | ibex | Spain |
| 15 | Jakarta Stock Exchange | jkse | Indonesia |
| 16 | Stock Market index of South Korea | ks11 | South Korea |
| 17 | Italian Bourse | mibtel | Italy |
| 18 | Bolsa Mexicana de Valores | mxx | Mexico |
| 19 | Tokyo Stock Exchange | nikkei | Japan |
| 20 | NASDAQ | ndx | USA |
| 21 | New York Stock Exchange | nya | USA |
| 22 | Stock exchange of Portugal | psi20 | Portugal |
| 23 | Shanghai Stock Exchange | ssec | China |
| 24 | Swiss Market Index | ssmi | Switzerland |
| 25 | Straits Times Index | sti | Singapore |

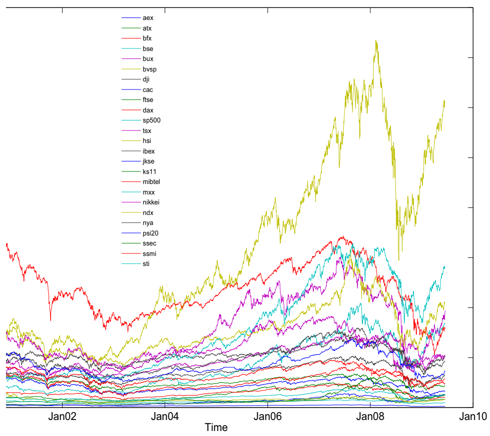


Figure 1. Time series for the twenty five indices from January 2000, up to December 2009.

We first compute the correlations among the twenty five stock markets obtained a $S \times S$ matrix and then apply MDS. In this representation, points represent the stock markets.

In order to reveal possible relationships between the market stocks index the MDS technique is used. In this perspective several MDS criteria are tested. The Sammon criterion revealed good results and is adopted in this work [16]. For this purpose we calculate $25 \times 25$ matrix **M1** based on the squared cosine coefficient $\xi(i, j)$, that provides a measurement of the similarity between two indices and is defined in equation (3). In matrix **M1** each cell represents the squared cosine correlation between a pair of indices, $i, j = 1, \cdots, S$.

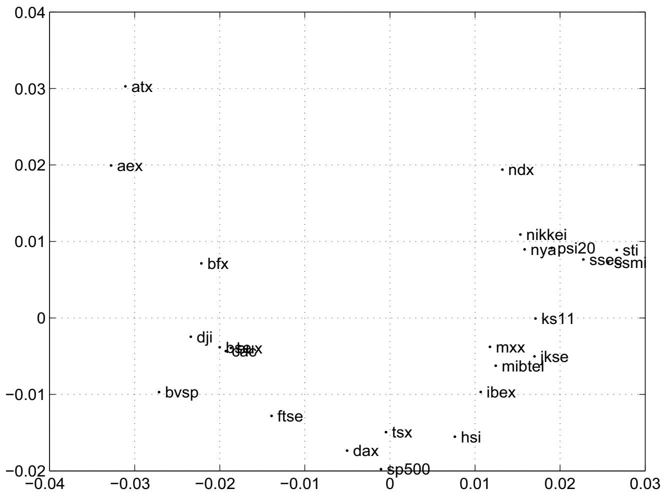


Figure 2. Two dimensional MDS graph for the twenty five indices using squared cosine correlation correlation, according (3).

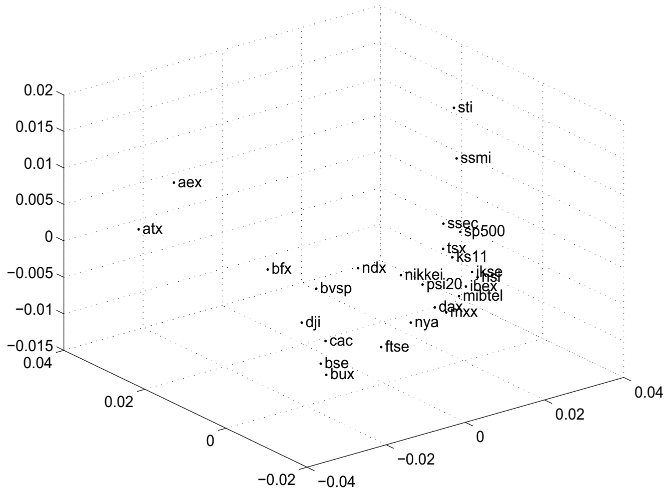


Figure 3. Three dimensional MDS graph for the twenty five indices using squared cosine correlation correlation, according (3).

$$\xi(i, j) = \left( \frac{\sum_{t=1}^{n} x_i(t) \cdot x_j(t)}{\sqrt{\sum_{t=1}^{n} (x_i(t))^2 \cdot \sum_{t=1}^{n} (x_j(t))^2}} \right)^2 \tag{3}$$

Figures 2 and 3, show the 2D and 3D locus of each index positioning in the perspective of expression (3), respectively. Figure 4 depicts the stress as function of the dimension of the representation space, revealing that a three dimensional space describes with reasonable accuracy the "map" of the twenty five signal indices. Moreover, the resulting Sheppard plot, represented in figure 5, shows that a good distribution of points around the 45 degree line is obtained [17].

There are several empirical conclusions one can draw from the graphs in figures 2 and 5, and we will mention just a few here. We can clearly observe that there seem to emerge

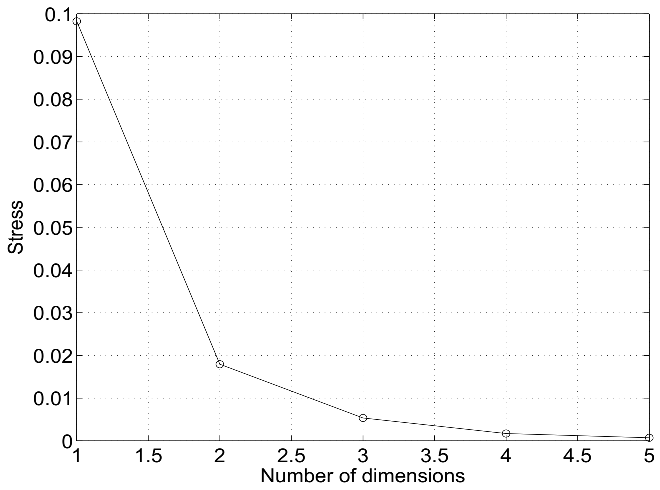


Figure 4. Stress plot of MDS representation of the twenty five indices *vs* number of dimension using squared cosine correlation correlation, according (3).

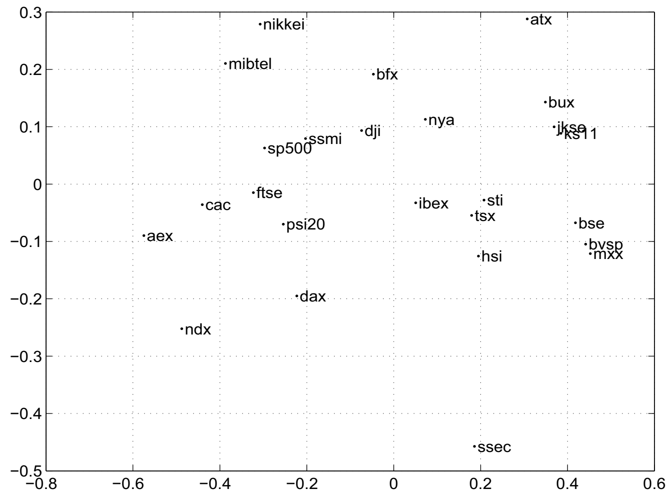


Figure 6. Two dimensional MDS graph for the twenty five indices using Pearson correlation, according (4).

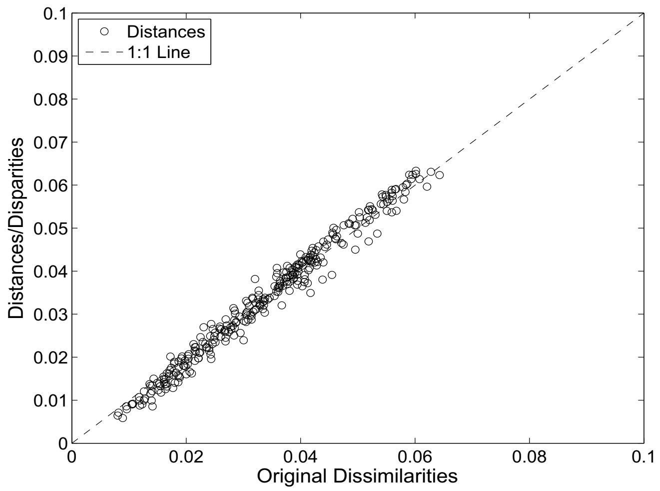


Figure 5. Shepard plot for MDS with a three dimensional representation of the twenty five indices using squared cosine correlation correlation, according (3).

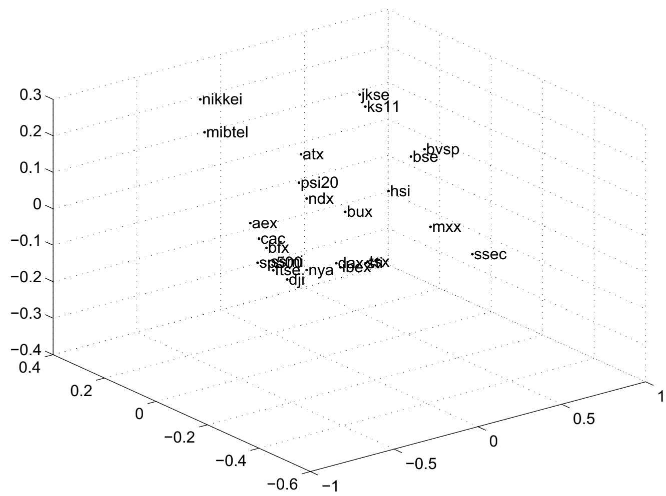


Figure 7. Three dimensional MDS graph for the twenty five indices using Pearson correlation, according (4).

clusters, which show similar behavior [18]. Hence, there does not seem to be a single world market, but perhaps there are several important regional markets. This last observation would match with standard financial theory which tells us that higher (lower) volatility corresponds with higher (lower) returns. Indeed, if this would be the case, one would expect to see similar patterns over time across returns and volatility.

### B. MDS analysis based on Pearson correlation

In this subsection we apply the MDS method using as "similarity measure" the values of Pearson correlation's coefficient correlation $\rho(i,j)$ defined in equation (4).

$$\rho(i,j)=\frac{\sum_{t=1}^{n}\left(x_i(t)-\overline{x_i}\right)\cdot\left(x_j(t)-\overline{x_j}\right)}{\sqrt{\sum_{t=1}^{n}\left(x_i(t)-\overline{x_i}\right)^2}\cdot\sqrt{\sum_{t=1}^{n}\left(x_j(t)-\overline{x_j}\right)^2}} \tag{4}$$

Figures 6 and 7, show the 2D and 3D locus of each index positioning in the perspective of expression (4), respectively. Figure 8 depicts the stress as function of the dimension of the representation space, revealing that a three dimensional space describes with reasonable accuracy the "map" of the twenty five signal indices. Moreover, the resulting Sheppard plot, represented in figure 9, shows that a good distribution of points around the 45 degree line is obtained [17].

### C. MDS analysis based on Kendall-tau correlation

In this subsection we apply the MDS method using as "similarity measure" the values of Kendall-tau correlation

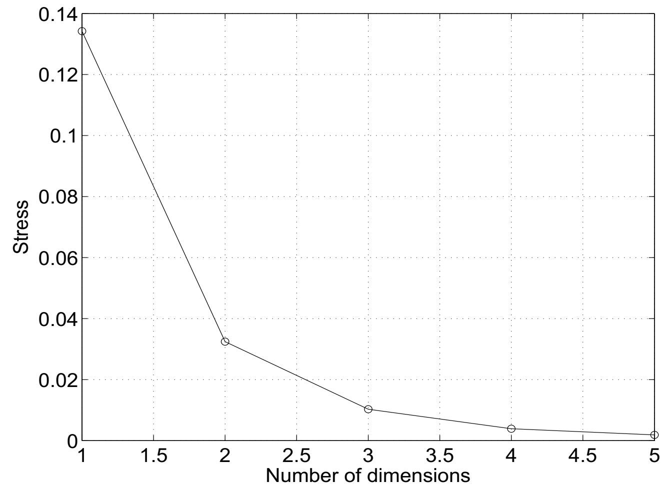


Figure 8. Stress plot of MDS representation of the twenty five indices *vs* number of dimension using Pearson correlation, according (4).

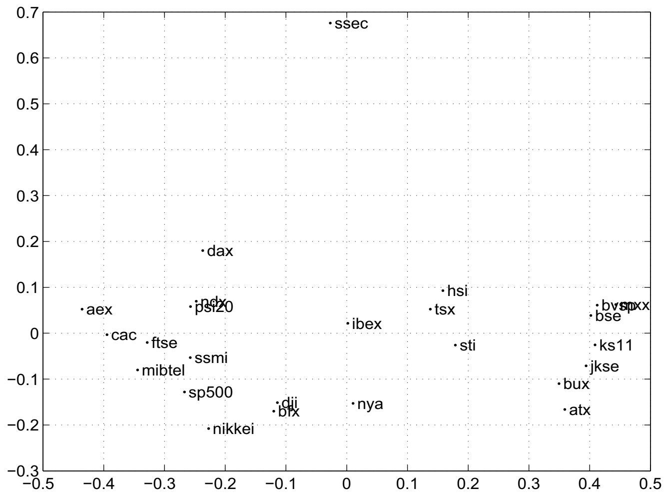


Figure 10. Two dimensional MDS graph for the twenty five indices using Kendall-tau correlation, according (5).

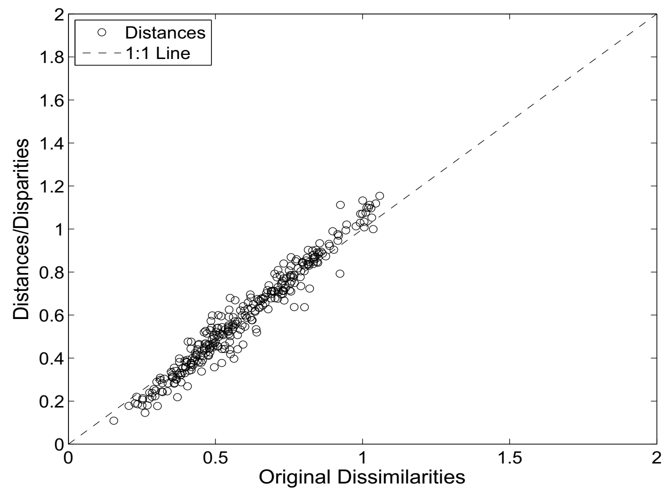


Figure 9. Shepard plot for MDS with a three dimensional representation of the twenty five indices using Pearson correlation, according (4).

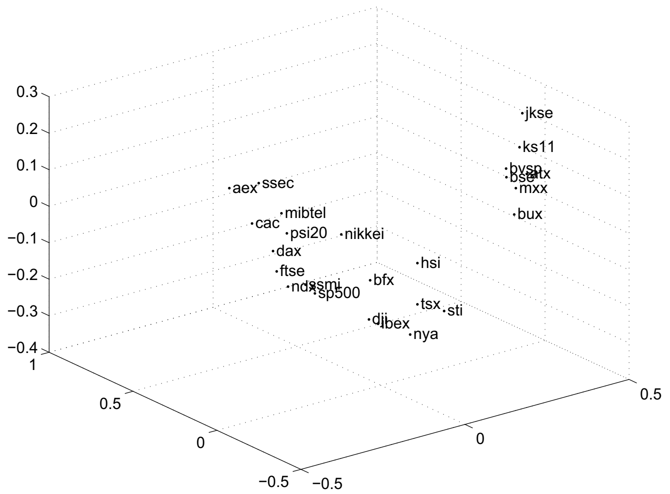


Figure 11. Three dimensional MDS graph for the twenty five indices using Kendall-tau correlation, according (5).

$\tau(i, j)$ defined in equation (5). Each pair of data points $(x_i, x_j)$ is classified as *concordant* (C), *discordant* (D) or *tied* (T). The pair is *concordant* if both variables increase or both variables decrease. The pair is *discordant* if one variable increases while the other one decreases. The pair says *tied* when one or both variables stays constant.

Writing $C$, $D$ and $T$ for the number of concordant, discordant and tied pairs, Kendall's coefficient is given by:

$$\tau(i, j) = \frac{C - D}{\frac{N}{2}(N - 1)} \tag{5}$$

where $N = C + D + T$ is the total number of pairs.

The idea is that *concordant* pairs suggest an increasing relationship, while *discordant* pairs suggest a decreasing relationship. Kendall's $\tau$ is just the proportion of concordant pairs minus the proportion of discordant pairs.

Figures 10 and 11, show the 2D and 3D locus of each index positioning in the perspective of expression (5), respectively. Figure 12 depicts the stress as function of the dimension of the representation space, revealing that a three dimensional space describes with reasonable accuracy the "map" of the twenty five signal indices. Moreover, the resulting Sheppard plot, represented in figure 13, shows that a good distribution of points around the 45 degree line is obtained [17].

### *D. MDS analysis based on Spearman correlation*

In this subsection we apply the MDS method using as "similarity measure" the values of Spearman correlation $\sigma(i, j)$ defined in equation (6).

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. Ranking both sets of data $x_i$ and $x_j$, from the highest to the lowest we have the correspondent ranks $xr_i$ and $xr_j$.

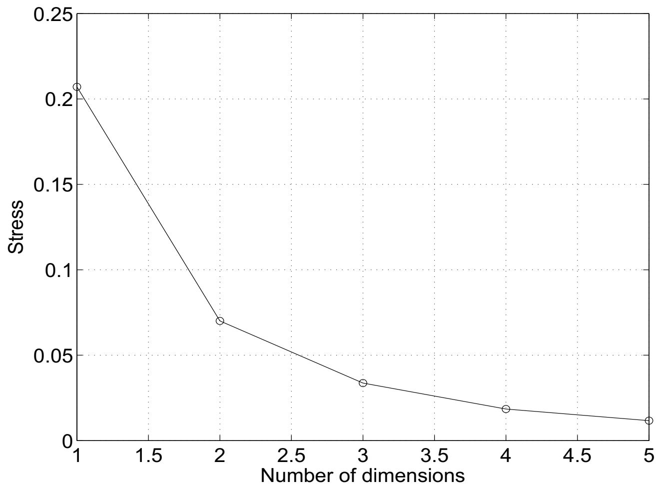


Figure 12. Stress plot of MDS representation of the twenty five indices *vs* number of dimension using Kendall-tau correlation, according (5).

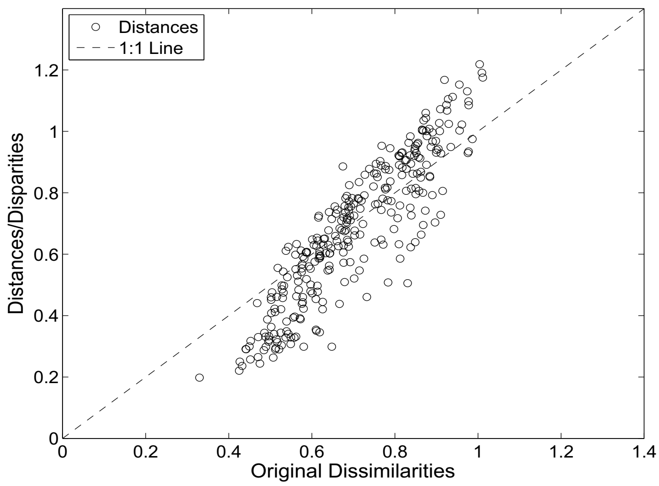


Figure 13. Shepard plot for MDS with a three dimensional representation of the twenty five indices using Kendall-tau correlation, according (5).

Tied values are assigned a rank equal to the average of their positions in the ascending order of the values.

The Spearman correlation $\sigma$ is computed from these:

$$\sigma(i,j)=\frac{\sum_{k=1}^{n}\left(xr_i(k)-\overline{xr_i}\right)\cdot\left(xr_j(k)-\overline{xr_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(xr_i(k)-\overline{xr_i}\right)^2\cdot\sum_{k=1}^{n}\left(xr_j(k)-\overline{xr_j}\right)^2}} \tag{6}$$

Figures 14 and 15, show the 2D and 3D locus of each index positioning in the perspective of expression (6), respectively. Figure 16 depicts the stress as function of the dimension of the representation space, showing that a three dimensional space “maps” adequately describes with reasonable accuracy the “map” of the twenty five signal indices. Moreover, the resulting Sheppard plot, represented in figure 17, shows good distribution of points around the 45 degree line.

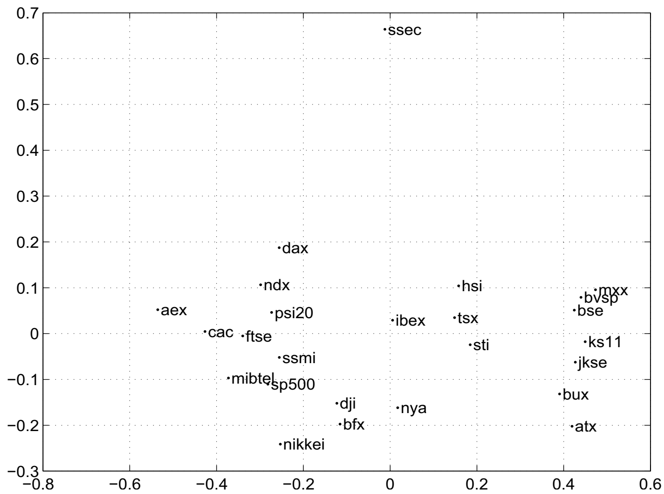


Figure 14. Two dimensional MDS graph for the twenty five indices using Spearman correlation, according (6).

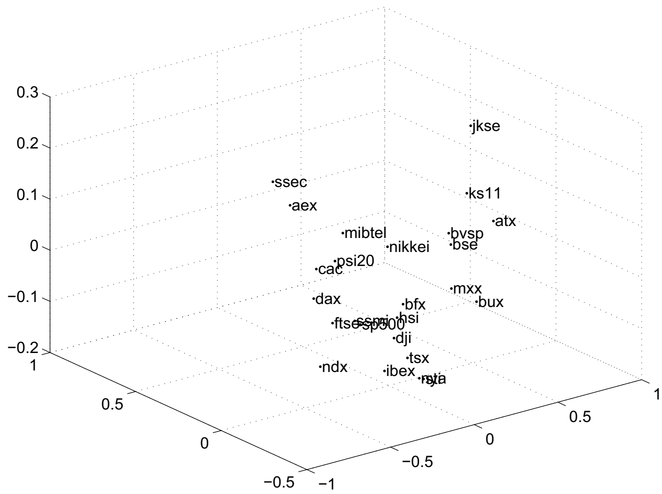


Figure 15. Three dimensional MDS graph for the twenty five indices using Spearman correlation, according (6).

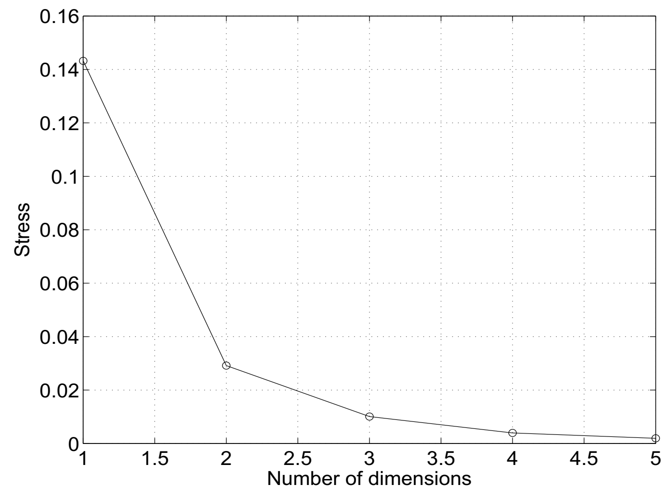


Figure 16. Stress plot of MDS representation of the twenty five indices *vs* number of dimension using Spearman correlation, according (6).

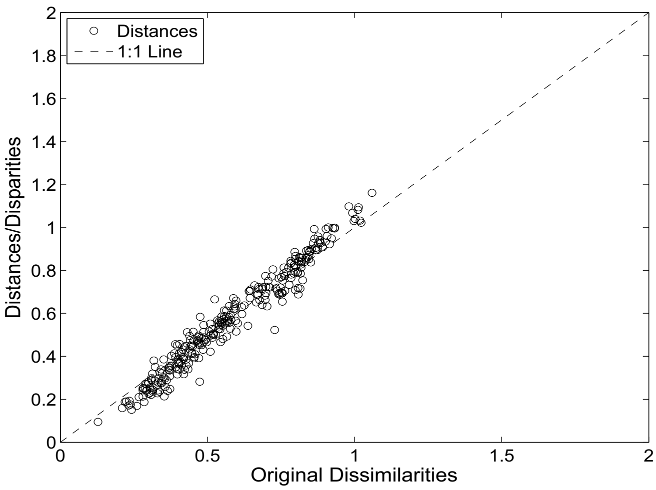


Figure 17. Shepard plot for MDS with a three dimensional representation of the twenty five indices using Spearman correlation, according (6).

## IV. Conclusion

In this paper, we proposed simple graphical tools to visualize time-varying correlations between stock market behavior. We illustrated our MDS-based method daily close values of fifteen stock markets. There are several issues relevant for further research. A first issue concerns applying our method to alternative data sets, with perhaps different sampling frequencies or returns and absolute returns, to see how informative the method can be in other cases. A second issue concerns taking the graphical evidence seriously and incorporating it in an econometric time series model to see if it can improve empirical specification strategies.

## References


[1] P. V., G. P. R. B., A. L.A.N., and S. H.E, "Econophysics: financial time series from a statistical physics point of view," *Physica A*, vol. 279, pp. 443–456, 2000.

[2] R. Nigmatullin, "Universal distribution function for the strongly-correlated fluctuations: General way for description of different random sequences," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 3, pp. 637–647, 2010.

[3] R. V. Mendes and M. J. Oliveira, "A data-reconstructed fractional volatility model," *Economics: The Open-Access, Open-Assessment E-Journal*, vol. 2, no. 2008-22, 2008. [Online]. Available: http://www.economics-ejournal.org/economics/discussionpapers/2008-22

[4] R. V. Mendes, "A fractional calculus interpretation of the fractional volatility model," *Nonlinear Dynamics*, vol. 55, no. 4, pp. 395–399, 2009.

[5] J. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," *Psychometrika*, vol. 29, no. 1, pp. 1–27, 1964. [Online]. Available: http://dx.doi.org/10.1007/BF02289565

[6] I. Borg and P. Groenen, *Modern Multidimensional Scaling:Theory and Applications*. New York: Springer, 2005.

[7] T. Cox and M. Cox. New York: Chapman & HallCrc, 2001.

[8] J. Kruskal and M. Wish, *Multidimensional Scaling*. Newbury Park, CA: Sage Publications, Inc., 1978.

[9] J. Woelfel and G. A. Barnett, "Multidimensional scaling in riemann space," *Quality and Quantity*, vol. 16, no. 6, pp. 469–491, 1982.

[10] J. O. Ramsay, "Some small sample results for maximum likelihood estimation in multidimensional scaling," *Psychometrika*, vol. 45, no. 1, pp. 139–144, 1980.

[11] S. Nirenberg and P. E. Latham, "Decoding neuronal spike trains: how important are correlations?" *Proc. National Academy of Sciences*, vol. 100, no. 12, pp. 7348–7353, 2003.

[12] F. B. Duarte, J. T. Machado, and G. M. Duarte, "Dynamics of the dow jones and the nasdaq stock indexes," *Nonlinear Dynamics*, vol. 61, no. 4, pp. 691–705, 2010.

[13] A. Buja, D. Swayne, M. Littman, N. Dean, H.Hofmann, and L.Chen, "Data visualization with multidimensional scaling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 85, pp. 444–472, 2008.

[14] "http://finance.yahoo.com."

[15] "http://www.bse.hu."

[16] B. Ahrens, "Distance in spatial interpolation of daily rain gauge data," *Hydrology and Earth System Sciences*, no. 10, pp. 197–208, 2006.

[17] J. T. Machado, F. B. Duarte, and G. M. Duarte, "Multidimensional scaling analysis of stock market values," in *Proceedings of 3rd Conference on Nonlinear Science and Complexity*, D. Baleanu, Ed. Cankaya University, July 2010.

[18] J. T. Machado, G. M. Duarte, and F. B. Duarte, "Identifying economic periods and crisis with the multidimensional scaling," *Nonlinear Dynamics*, DOI 10.1007/s11071-010-9823-2,ISSN: 0924-090X.