# Multidimensional Scaling Analysis of Stock Market Values

**J. Tenreiro Machado** * **Fernando B. Duarte** ** **Gonçalo Monteiro Duarte** ***

* *Dept. of Electrical Engineering, Institute of Engineering, Porto, Portugal, (e-mail: jtm@isep.ipp.pt)*
** *Dept. of Mathematics, School of Technology, Viseu, Portugal, (e-mail: fduarte@estv.ipv.pt).*
*** *e-mail: monteiro.duarte@gmail.com*

**Abstract:** We propose a graphical method to visualize possible time-varying correlations between fifteen stock market values. The method is useful for observing stable or emerging clusters of stock markets with similar behavior. The graphs, originated from applying multidimensional scaling techniques (MDS), may also guide the construction of multivariate econometric models.

*Keywords:* Multidimensional scaling; Stock market daily values; Time-varying correlation.

## 1. INTRODUCTION

Economical indexes 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 analyses possible correlations/similarities in international stock markets. Its results are expected to guide the design of statistical models aiming to test hypotheses of interest. Ultimately, the method can even lead to the postulation of new hypotheses. 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 (Plerou et al., 2000; Nigmatullin, 2010).

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 fifteen 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.

## 2. MULTIDIMENSIONAL SCALING

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 judgements (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 (Borg and Groenen, 2005). 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 analyse 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 theorectically needed to a perfectly fit is reached (i.e. $N-1$ dimensions for $N$ objects) (Cox and Cox, 2001; Kruskal and Wish, 1978; Woelfel and Barnett, 1982; Ramsay, 1980).

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 (Nirenberg and Latham, 2003). 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 condi-

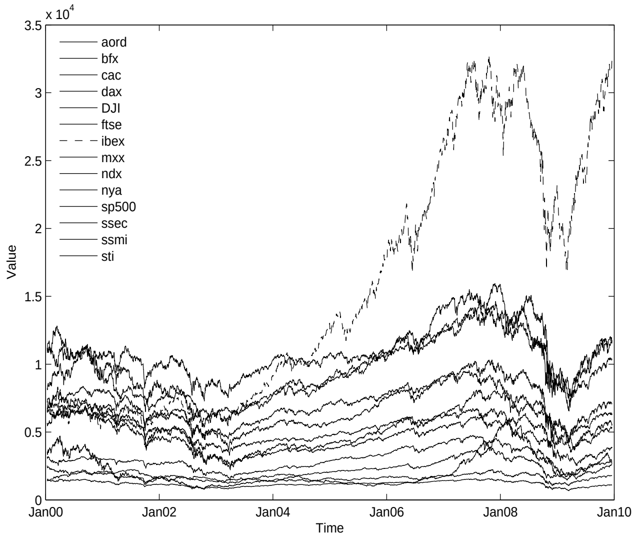


Fig. 1. Time series for the fifteen indexes from January 2000, up to December 2009.

tional 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 (Buja et al., 2008).

## 3. ANALYSIS OF STOCKS MARKETS

In this section we study numerically the fifteen selected stock markets, including six American markets, six European markets and three Asian/Pacific markets.

Our data consist of the $n$ daily close values of $S = 15$ stock markets from January 2, 2000, up to December 31, 2009, to be denoted as $x_i(t), 1 \leq t \leq n$, $i = 1, \cdots, S$. The stock markets are listed in Table 1.

Table 1. Fifteen stock markets

| $i$ | Stock market index | Abbreviation | Country |
|---|---|---|---|
| 1 | All Ordinaries | aord | Australia |
| 2 | EURONEXT BEL-20 | bfx | Belgium |
| 3 | Cotation Assistée en Continu | cac | France |
| 4 | Deutscher Aktien Index | dax | German |
| 5 | Dow Jones Industrial | dji | USA |
| 6 | Footsie | ftse | UK |
| 7 | Iberia Index | ibex | Spain |
| 8 | Bolsa Mexicana de Valores | mxx | Mexico |
| 9 | NASDAQ | ndx | USA |
| 10 | New York Stock Exchange | nya | USA |
| 11 | Standard & Poor's | sp500 | USA |
| 12 | Shanghai Stock Exchange | ssec | China |
| 13 | Swiss Market Index | ssmi | Swiss |
| 14 | Straits Times Index | sti | Singapore |
| 15 | Toronto Stock Exchange | tsx | Canada |

The data are obtained from data provided by Yahoo Finance web site http://finance.yahoo.com (2010), and they measure indexes in local currencies.

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

The section is organized in two subsections, the first adopts an analysis based on the correlation of the time evolution and the second adopts a metrics based on histogram distances.

### *3.1 MDS analysis based on time correlation*

In this subsection, we apply the MDS method described in Section 2 to the time correlation of the selected stock markets.

For the fifteen markets, we consider the time correlations between the daily close values. We first compute the correlations among the fifteen 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 (Lima and Machado, 2009; Ahrens, 2006). For this purpose we calculate $15 \times 15$ matrix $\mathbf{M}$ based on a correlation coefficient $c(i, j)$, that provides a measurement of the similarity between two indexes and is defined in equation (1). In matrix $\mathbf{M}$ each cell represents the time correlation between a pair of indexes:

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

$i, j = 1, \cdots, S$. Figures 2 and 3, show the 2D and 3D locus of each index positioning in the perspective of expression (1), respectively. Figure 4 depicts the stress as function of the dimension of the representation space, revealing that a three dimensional space describe a with reasonable accuracy the "map" of the fifteen signal indexes. Moreover, the resulting Sheppard plot, represented in figure 5, shows that a good distribution of points around the 45 degree line is obtained.

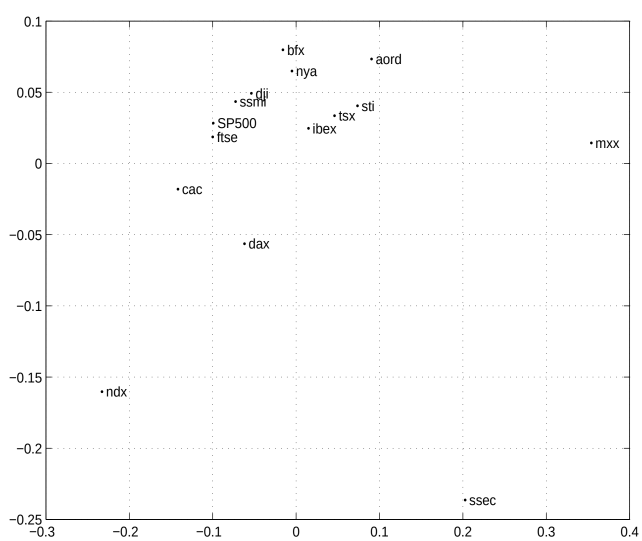


Fig. 2. Two dimensional MDS graph for the fifteen indexes using time correlation.

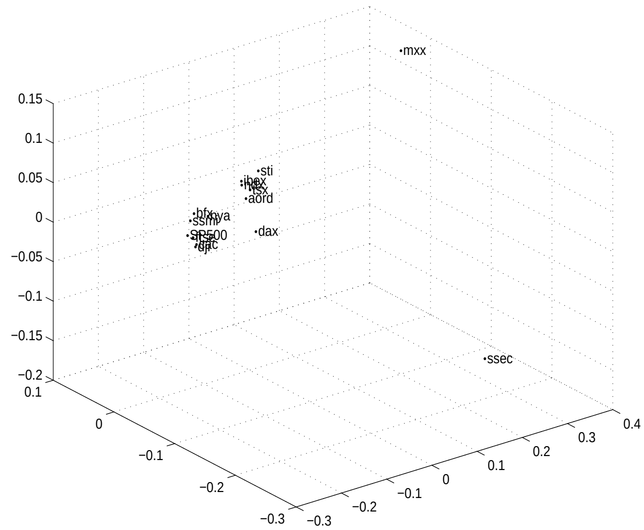


Fig. 3. Three dimensional MDS graph for the fifteen indexes using time correlation.

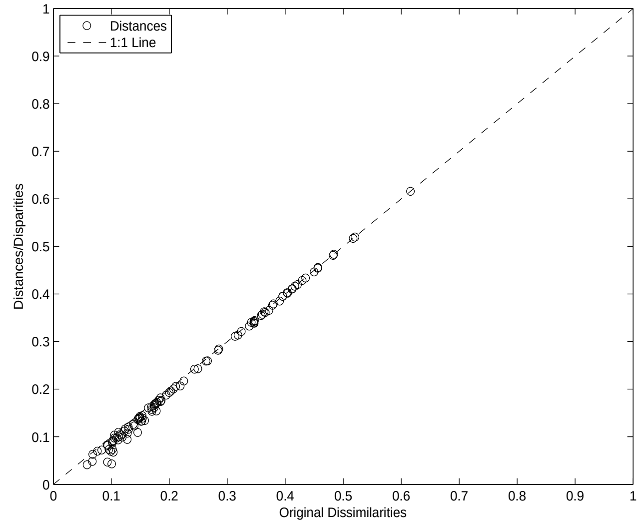


Fig. 5. Shepard plot for MDS with a three dimensional representation of the fifteen indexes using time correlation.

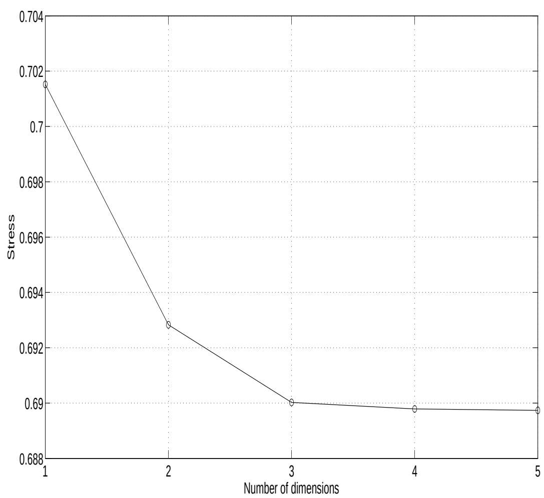


Fig. 4. Stress plot of MDS representation of the fifteen indexes *vs* number of dimension using time correlation.

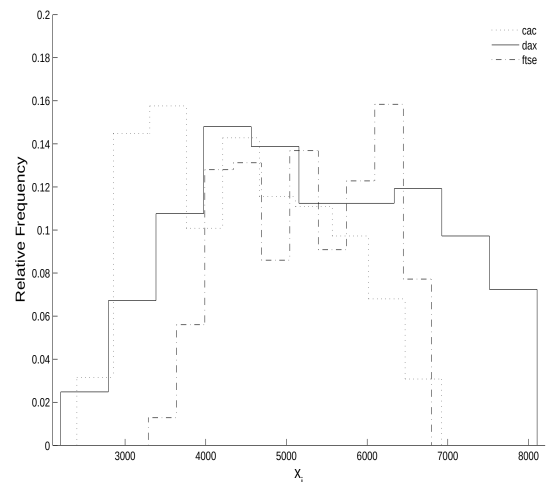


Fig. 6. Histogram for the CAC, DAX and FTSE indexes over all time.

There are several empirical conclusions one can draw from the graphs in figures 2 and 3, and we will mention just a few here. We can clearly observe that there seem to emerge clusters, which show similar behavior. 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.

### *3.2 MDS analysis based on histogram*

For each of the fifteen indexes we draw the corresponding histogram of relative frequency and we calculate statistical descriptive parameters like the arithmetic mean ($\mu_i$), the standard deviation ($\sigma_i$) and the Pearson's Kurtosis coefficient $\gamma_i$.

Figures 6 and 7 decipt the histograms of the CAC, DAX, DJI, FTSE, NDX and NYA indexes. The values of the statistical descriptive parameters are listed in Table 2.

For all the fifteen indexes we calculate the "histogram's distance" (Serratosa and Sanroma, 2008; Sierra et al., 2009), $d_1$ and $d_2$ using the equations:

$$d_1(i,j) = \sqrt{\frac{(\mu_i - \mu_j)^2}{\mu_i^2 + \mu_j^2} + \frac{(\sigma_i - \sigma_j)^2}{\sigma_i^2 + \sigma_j^2}} \quad (2a)$$

$$d_2(i,j) = \sqrt{\frac{(\mu_i - \mu_j)^2}{\mu_i^2 + \mu_j^2} + \frac{(\sigma_i - \sigma_j)^2}{\sigma_i^2 + \sigma_j^2} + \frac{(\gamma_i - \gamma_j)^2}{\gamma_i^2 + \gamma_j^2}} \quad (2b)$$

where $i, j = 1, \cdots, S$.

Figures 8-11, show the 2D and 3D locus of each index positioning in the perspective of the expressions (2a) and

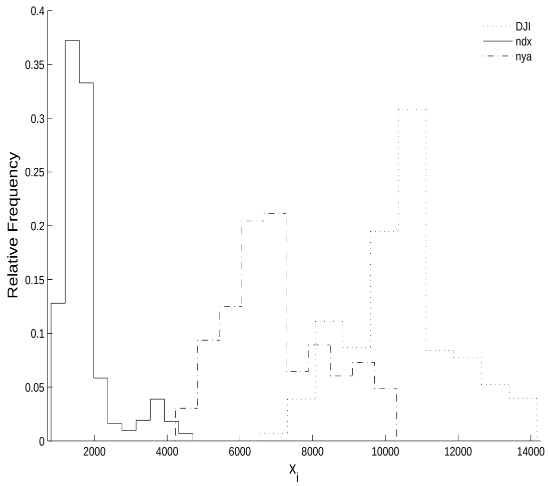


Fig. 7. Histogram for the DJI, NDX and NYA indexes over all time.

Table 2. Statistical Descriptive Parameters

| $i$ | $\mu_i$ | $\sigma_i$ | $\gamma_i$ |
|---|---|---|---|
| 1 | 4082.52 | 1074.17 | -0.51 |
| 2 | 2956.11 | 788.96 | -0.67 |
| 3 | 4475.70 | 1071.94 | -1.04 |
| 4 | 5324.72 | 1440.27 | -1.00 |
| 5 | 10472.98 | 1454.40 | -0.11 |
| 6 | 5248.86 | 871.57 | -1.24 |
| 7 | 10042.82 | 2583.65 | -0.72 |
| 8 | 15372.58 | 9168.40 | -1.27 |
| 9 | 1753.80 | 701.35 | 4.05 |
| 10 | 7034.10 | 1404.76 | -0.58 |
| 11 | 1187.55 | 198.44 | -0.85 |
| 12 | 2079.76 | 1031.71 | 2.96 |
| 13 | 6689.10 | 1337.82 | -0.95 |
| 14 | 2179.07 | 615.19 | -0.22 |
| 15 | 9789.64 | 2360.58 | -0.95 |

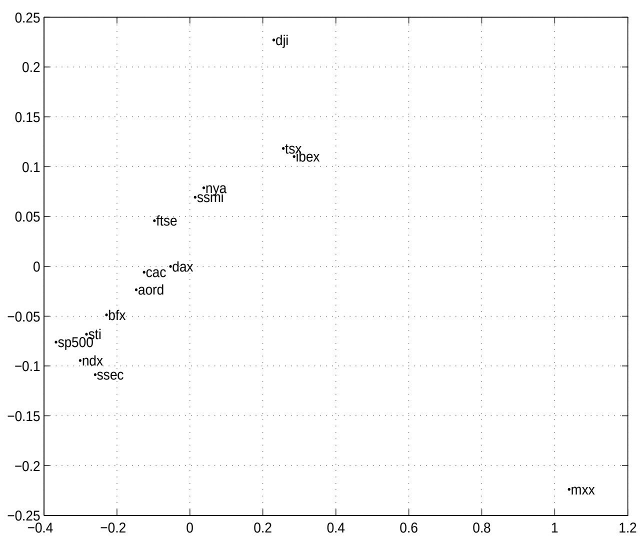


Fig. 8. Two dimensional MDS graph for the fifteen indexes using histogram's distance $d_1$.

(2b), respectively demonstrating differences between the corresponding MDS plots.

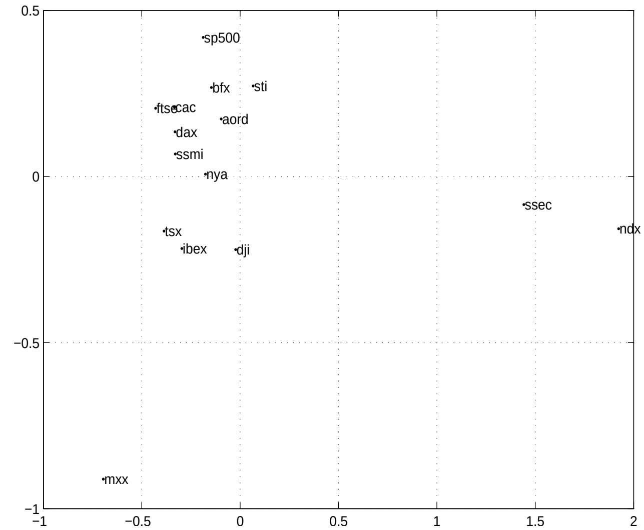


Fig. 9. Two dimensional MDS graph for the fifteen indexes using histogram's distance $d_2$.

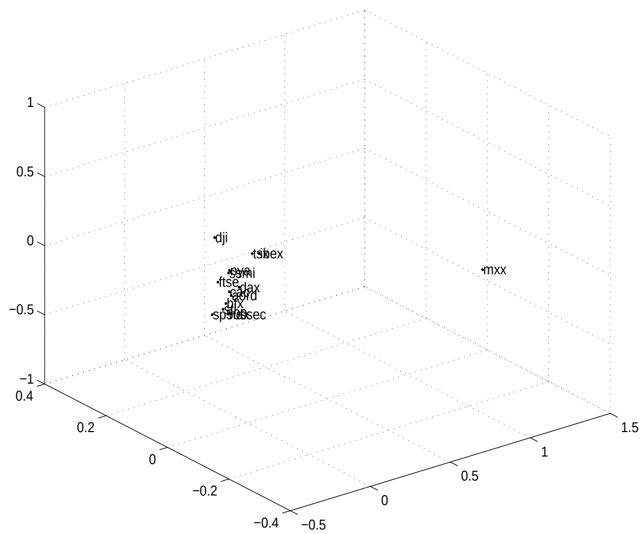


Fig. 10. Three dimensional MDS graph for the fifteen indexes using histogram's distance $d_1$.

Figures 12-13 depict the stress as function of the dimension of the representation space based on $d_1$ and $d_2$ distances, revealing that a three dimensional space describe with reasonable accuracy the "map" of the fifteen signal indexes. Moreover, the resulting Sheppard plot, represented in figures 14-15, show that a good distribution of points around the 45 degree line is obtained for the two indices.

Curiously in the chart corresponding to the MDS based on correlation (figure 2) we can see an V shape with the NDX index at the vertex, and the BFX and AORD at the corners. The MXX and the SSEC indexes are out of the angle form. However in the chart corresponding to the MDS based on the histogram distance (figures 8 and 9) such an angle form cannot be found. Instead $d_1$ leads to a long "S" curve having the DJI and the SSEC indexes as extremes emerges can be observed in figure 8. On the other hand, $d_2$ produces the map of figure 9 where the SSEC, NDX and MXX are far apart from the rest of the points similarly to what occurs in the map of figure 2.

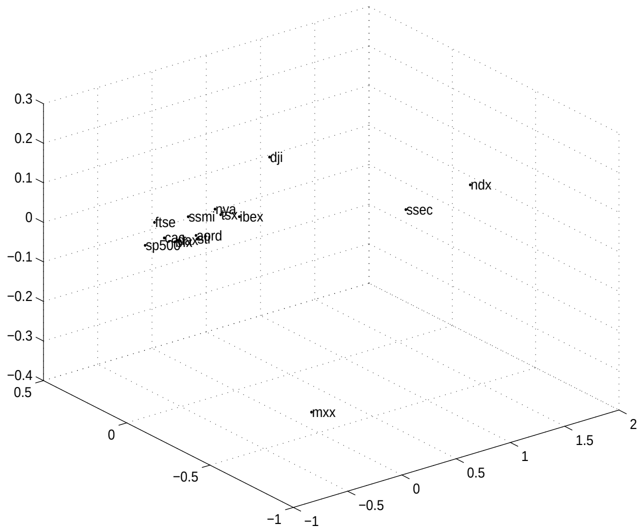


Fig. 11. Three dimensional MDS graph for the fifteen indexes using histogram's distance $d_2$.

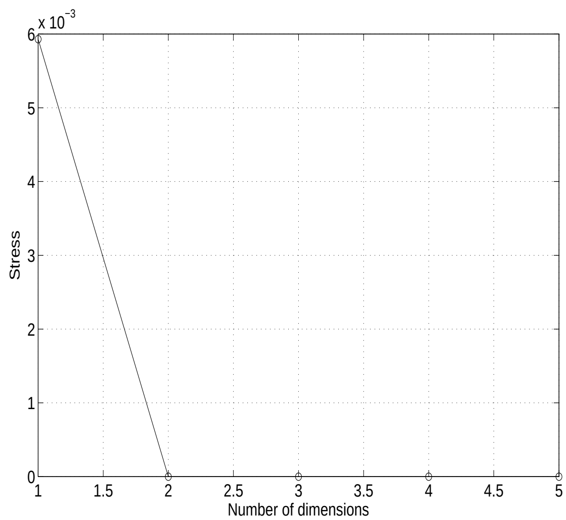


Fig. 12. Stress plot of MDS representation of the fifteen indexes *vs* number of dimension, using histogram's distance $d_1$.

It is interesting to note that in all cases the MXX index behaves differently from the other (i.e., is not part of the shapes and regularities formed). Perhaps this may explained by the fact that Mexico was less affected by the *dot.com* crisis in the beginning of the period under study, since then it was strongly emerging from its own *Mexican Peso Crisis*.

## 4. 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

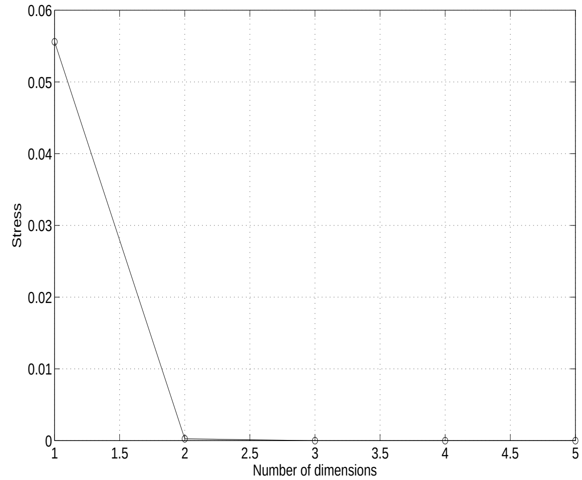


Fig. 13. Stress plot of MDS representation of the fifteen indexes *vs* number of dimension, using histogram's distance $d_2$.

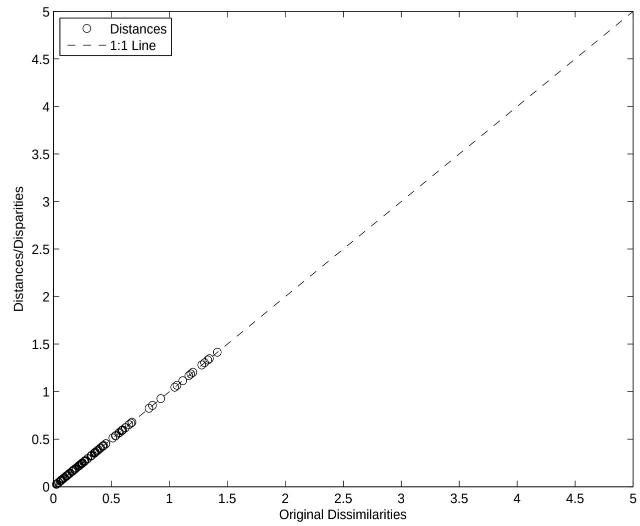


Fig. 14. Shepard plot for MDS with a three dimensional representation of the fifteen indexes, using histogram's distance $d_1$.

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


Ahrens, B. (2006). Distance in spatial interpolation of daily rain gauge data. *Hydrology and Earth System Sciences*, (10), 197–208.

Borg, I. and Groenen, P. (2005). *Modern Multidimensional Scaling:Theory and Applications*. Springer, New York.

Buja, A., Swayne, D., Littman, M., Dean, N., H.Hofmann, and L.Chen (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, 17(85), 444–472.

Cox, T. and Cox, M. (2001). *Multidimensional Scaling*. Chapman & HallCrc, New York.

http://finance.yahoo.com (2010).

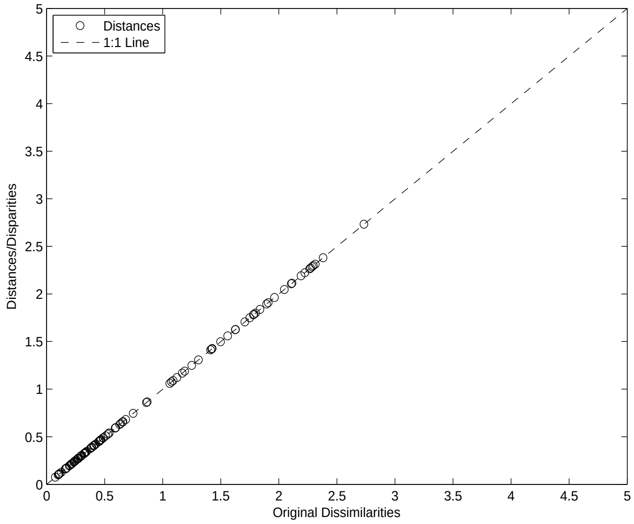


Fig. 15. Shepard plot for MDS with a three dimensional representation of the fifteen indexes, using histogram's distance $d_2$.


Kruskal, J. and Wish, M. (1978). *Multidimensional Scaling*. Sage Publications, Inc., Newbury Park, CA.

Lima, M. and Machado, J.A.T. (2009). A sensor classification scheme of robotic manipulators using multidimensional scaling technique. In *Proccedings of theInternational Symposium on Computational Intelligence for Engineering Systems*. Porto, Portugal.

Nigmatullin, R. (2010). Universal distribution function for the strongly-correlated fluctuations: General way for description of different random sequences. *Communications in Nonlinear Science and Numerical Simulation*, 15(3), 637–647.

Nirenberg, S. and Latham, P.E. (2003). Decoding neuronal spike trains: how important are correlations? *Proc. National Academy of Sciences*, 100(12), 7348–7353.

Plerou, V., Gopikrishnan, P., Rosenow, B., Amaral, L., and Stanley, H. (2000). Econophysics: financial time series from a statistical physics point of view. *Physica A*, 279, 443–456.

Ramsay, J.O. (1980). Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika*, 45(1), 139–144.

Serratosa, F. and Sanroma, G.and Sanfeliu, A. (2008). A new algorithm to compute the distance between multidimensional histograms. *Lecture Notes in Computer Science*, 4756(1), 115–123.

Sierra, B., Lazkano, E., Jauregi, E., and Irigoien, I. (2009). Histogram distance-based bayesian network structure learning: A supervised classification specific approach. *Decis. Support Syst.*, 48(1), 180–190. doi: http://dx.doi.org/10.1016/j.dss.2009.07.010.

Woelfel, J. and Barnett, G.A. (1982). Multidimensional scaling in riemann space. *Quality and Quantity*, 16(6), 469–491.