Mapping of Stock Exchanges: Contagion from a new perspective

Pablo García Estévez[a1], Salvador Roji Ferrari[b], Teresa Corzo Santamaría[c].

a CUNEF, Madrid

b Universidad Complutense de Madrid

c Universidad Pontificia Comillas, ICAI-ICADE

Pablo García Estévez, PhD, Master, Associate Professor at CUNEF, Madrid. Published 9 books in fields related to Finance, Markets and Mergers. Research published in Intelligent System in Accounting, Finance and Management, Journal of Finance Education and Spanish Economics Review. Currently teaching Corporate Finance at the bachelor and master's level

Salvador Roji Ferrari, PhD, Associate Professor, Universidad Complutense de Madrid Published 5 books in fields related to Finance, Business Enterprise, and Complexity Theory. Currently teaching Corporate Finance and Multinational Financial Management at the bachelor and master ́s level

Teresa Corzo Santamaría, PhD, CFA, is Associate Professor at Universidad Pontificia Comillas, ICADE, Madrid. Interest areas are Behavioural Finance, International Finance Asset pricing and Portfolio Management. Research published in Journal of Behavioural Finance, Journal of Asset Management, Spanish Economics Review, Economic Notes and Universia Business Review. Currently teaching Derivatives and Portfolio Management at the bachelor and master´s level.

[1] Corresponding author at: Colegio Universitario de Estudios Financieros (CUNEF). C/ Leonardo Prieto Castro 2, 28040 Madrid. (Spain) Tfno: +34 91 448 08 92
Universidad Complutense. Campus de Somosaguas. Pozuelo de Alarcón 28223 Madrid (Spain)
Universidad Pontificia de Comillas. ICADE. Alberto Aguilera, 23 - 28015 Madrid (Spain)
Email address: pgestevez@cunef.edu (García Estévez), srojifer@ccee.ucm.es (Roji Ferrari), mcorzo@cee.upcomillas.es (Corzo Santamaría)

Mapping of Stock Exchanges: Contagion from a new perspective

Abstract

This paper addresses the contagion problem between stock exchanges and uses self-organizing maps (SOMs) to develop a two-dimensional map based on five market variables from 48 stock exchanges around the world for the 2000-2012 period. The study addresses markets worldwide that have different levels of economic development. Using five market variables, we group stock exchanges into distinctive clusters and analyze their similarities, differences, and dynamics. The technique applied, artificial neural networks (ANNs), is a non-parametric visualization tool to study the dynamics of the markets worldwide following a non-traditional perspective based on the specific formation of groups and their evolution. Each group is defined by the Euclidean distances between markets. There are certain common features, both economic and geographic, within the migrant and static exchanges, which offer new insights. By applying the standard contagion methodology to the resulting groups, we study the spread of the subprime crisis and find evidence of contagion, as well as a high interdependence between markets for the entire period under consideration.

Keywords: Clusters, International Financial Markets, Self-Organizing Neural Networks, Financial Contagion.

Mapeo de las bolsas de valores: El contagio desde una nueva perspectiva

Resumen

Este artículo aborda el problema del contagio entre bolsas y utiliza un mapa auto-organizado (SOM) para desarrollar un mapa de dos dimensiones basado en cinco variables del mercado pertenecientes a 48 mercados de todo el mundo durante el periodo 2000 al 2012. El estudio aborda mercados de países con diferentes niveles de desarrollo. Usando cinco variables hemos agrupado las bolsas en diferentes *clusters* y hemos analizado sus similitudes, diferencias y dinámicas. La técnica aplicada, redes neuronales artificiales (ANNs), constituye una herramienta de visualización no paramétrica para estudiar la dinámica de los mercados mundiales siguiendo una perspectiva no tradicional basada en la formación específica de grupos en su evolución. Cada grupo está definido por las distancias Euclídeas entre los mercados. Hay algunas características comunes, tanto económicas como geográficas, dentro de los mercados estables y de los que migran, que nos ofrecen nuevas ideas. Aplicando la metodología estándar de contagio a los grupos resultantes, estudiamos la influencia de la Crisis *Subprime* y encontramos evidencia de contagio, además de una alta interdependencia entre los mercados durante todo el periodo considerado.

Palabras Clave

Grupos, Mercados Financieros Internacionales. Redes Auto-Organizadas, Contagio Financiero

1 Introduction

During the last years, the study of the co-movements and contagion between stock exchanges in different as well as the same time zones has been a fruitful research area. Within this framework, knowledge regarding the ties between the stock exchanges, their similarities, and the level of connectivity may influence international dynamics. Using a novel approach, in this article we map stock exchanges covering most of the world, observe their relations, and learn from their dynamics within a context of periods of stability and volatility.

Based on five key characteristics that describe each stock exchange, we draw a detailed map of 48 financial markets around the globe using four clusters, which are defined as zones. We scan the exchanges in four non-overlapping economic periods to observe the evolution of the clusters during the 2000-2013 period and, specifically, the migration of exchanges between the clusters.

Several studies have analyzed contagion, correlations, and herd behavior in international capital markets (Quinn & Voth, 2008) using models such as cross-market correlations, ARCH and GARCH, co-integration, discriminant analysis, logit and probit, and multiple regressions. Our contribution to the current literature consists of exploring the contagion among groups of stock markets worldwide by using the SOM methodology. SOM methodology allows us to cluster exchanges based on information from several market-based variables and map them into a visual and intuitive two-dimensional chart. The standard methodologies (e.g. Forbes & Rigobon, 2002) used to study market contagion are based on correlations between two market variables. However, our approach considers that the contagion happens between sets of similar, homogeneous markets, being propagated unevenly to the rest of the groups.

Artificial Neural Networks (ANNs) are a set of non-linear data-driven self-adaptive approaches for modeling large quantities of raw and blurred databases and have already been successfully used for handwriting patterns (Basu, Bhattacharya & Kim, 2010), control (Hagan, Demuth & Jesús, 2002), and robotics (Jegede, Awodele, Ajayi & Ndong, 2007). ANNs have certain advantages over more traditional classification schemes. Mainly because of their adaptability, ANNs can identify and learn patterns between different complex data sets in an iterative manner, particularly when these sets are very noisy and largely imprecise, as is the case for the data available for many of the stock exchanges from emergent or less developed markets.

Once we have classified the 48 stock exchanges in our sample into groups, we determine a key macroeconomic event to explore the dynamics between groups. In accordance with Forbes and Rigobon (2002), we identify contagion as a significant increase in cross-market linkages after a shock.[1]

In this way, we combine the non-parametric and visualization tool self-organizing maps (SOMs) with the methodology developed by Forbes and Rigobon (2002). Our goal is twofold: to test if there was some level of contagion between stock markets during the crisis of 2008, and if positive, to test whether this contagion started in specific groups of exchanges.

The shock studied is the subprime crisis; instead of studying individual markets, we study groups of markets. Stock markets are economic leading indicators and can serve to assess a country's economic condition. Therefore, the interpretation of those shocks can be used as a forecasting tool. Frankel and Saravelos (2012) determine that variables related to financial markets may be of interest in crisis periods.

Because the contagion results are critically dependent on the way the classification into groups has been conducted, we relate our results to those of two other classification schemes as reference points to better understand the formation and dynamics of market clusters. Although their criteria, models, time span, and variables are very different, these classifiers can provide a sense of dimensionality and perspective to our proposal.

Our methodology is of interest to policy makers and market participants because of its ability to relate markets, and to assess the similarities. It also has a potential for assessing contagion, focusing on groups instead of individual stock-markets.

The remainder of this paper is organized as follows. In Section 2, the sample is explained: the stock exchanges, the time periods, and the market variables under study. Section 3 introduces the self-organizing maps (SOMs) model, contagion, and the empirical application. Section 4 analyzes and discusses the results in detail. Finally, Section 5 we present our conclusions. The references can be found at the end of the paper.

---

[1] As Forbes and Rigobon (2002) note, this is not a universally accepted definition of contagion.

2 Sample

The database used in this study is provided by the World Federation of Exchanges. The raw database included 34 variables and 146 markets from 2000 to 2013, and we use monthly data. However, not all markets provide data for every variable every month; therefore, we filtered the database to find the proper balance between the number of variables and the number of stock markets.

Five market variables turn out to be an adequate number of variables since adding more variables to the study significantly reduced the number of exchanges with total data available.[2] Since our method adapts the number of clusters to the number of observations, if we increase the number of variables but get fewer stock exchanges, the number of clusters is also significantly reduced. Besides, we find that, because of the high correlation shown by several variables, some of them become information-redundant when clustered.[3]

We divided the period of study into four non-overlapping periods of economic significance, and we look for the variables that encompassed the largest number of markets that meet the requirement of nearly complete data for at least three periods.

The resulting sample is comprised of 48 markets and five input variables (Table 1). The variables are: domestic market capitalization (DMC); electronic turnover (ET); foreign listings (LF); domestic listings (LD); and the number of companies newly listed through an IPO (NNC).

The reasons to use the five variables is threefold. First, as it has already been mentioned, they are the only ones with enough data to generate a significant number of clusters (see footnote # 2). Second, these variables are key to developing an understanding of the dynamics of the exchanges. The World Federation of Exchanges collects data from many

---

[2] For instance, if we include 13 variables, the number of observations is reduced to zero for the period 2000-2002; to four for the period 2003-2007, twelve for 2008-2009 and twenty-six for 2010-2013. The resulting clusters are two. Results are available upon request.

[3] The results of two different robustness studies considering seven and thirteen variables are available upon request. These studies show that results are persistent and robust to different model specifications.

developing countries and focuses on relevant variables that affect their emerging financial systems. Hence, their importance to the study of any systemic and contagion effect and, as a result, to our work. Third, these variables have extensively been employed extensively in many studies about global markets.

DMC is the total number of issued shares of domestic companies multiplied by their respective prices at a given time. It indicates the comprehensive value and size of the market. De la Torre, Gozzi and Schmukler (2007) use it as reference to assess the stock market development under globalization, while Levine and Schmukler (2006) study the liquidity and internationalization of stock markets, using this variable. ET is the number of shares traded through the exchange's electronic order book multiplied by their respective matching prices. ET is a proxy for number of transactions, commonly used to assess the bid-ask spreads in the foreign exchange market (Ding & Hiltrop, 2010). LD and LF are the companies listed on an exchange at the end of the period. We split them into domestic (LD) and foreign (LF). DL provides managers and policy makers with information about the effects of listings on stock prices. In general, new listings are associated with increases in value without increases in risk. (McConell, Dybevik, Haushalter & Lie, 1996). FL indicates the level of internationalization (Alexander, Eun & Janakiramanan, 1988; Pagano, Randl, Röell & Zechner, 2001). For Saudaragan and Biddle (1995), FL assesses factors that influence firms' choices of foreign stock exchanges. A company is considered domestic when it is incorporated in the same country where the exchange is located. LF is a key variable, which serves as a channel of contagion. ET and NNC provide information regarding the level of activity in the market. NNC, the underlying demand for initial public offerings, is often related to investor sentiment (Baker & Wurgler, 2007).[4]

These variables define and characterize each market, which allows us to draw comparisons and study similarities.

[Insert Table 1 about here]

[4] For more information regarding these variables, please consult http://www.world-exchanges.org/statistics/statistics-definitions.

To study the dynamics of global stock exchanges, we have divided the dataset according to key economic systemic events that have caused uneven and deep changes in many stock exchanges. To do so, we have used information from IMF quarterly reports (World Economic Outlook) and data from the International Stock Exchanges Association. We believe that an analysis of this length, with periods of stability, growth, and deep recession, can provide us the dimension needed to calibrate the dynamics of the markets. If the exchanges, on average, are partially efficient, we can derive a strong connection between certain key economic events and the resulting reaction in the markets. Apparently, due to low or negative correlations between regions, it is not easy to define and label the entire world during a specific period and constrain it to the same economic features. Nevertheless, if we focus on the variations in GDP around the world during these 14 years, we can separate the entire period into four clear-cut, non-overlapping segments along economic lines.

- 2000 - 2002, the dotcom bubble
- 2004 - 2007, the stable and recovery period
- 2008 -2009, the subprime mortgage crisis
- 2010 – 2013, the sovereign debt default crisis

The first segment covers the years 2000-2002, which is a weak, mixed, slowdown period that followed high growth rates around the world. In accordance with the financial press, we have labeled this period the dot.com bubble because, although it did not affect the entire world, a key financial event was the loss of $5 trillion in the capitalization of American stock exchanges from March 2000 to October 2002 and a sharp decline in global equity markets. The next segment from 2003 to 2007 is defined by the IMF as a global recovery and expansion period, with stable growth rates. The third segment, 2008-2009, has been labeled the subprime mortgage crisis. It represents the most dramatic financial shock since the Depression, triggering a deep global downturn. This crisis was eased by a rebound in certain areas of the world that compensated for the deep recession in the West. Finally, the last period, 2010-2013, represents the years of sluggish, uneven recovery due to the sovereign debt crisis and the uncertainty which followed it, with certain countries growing quickly, whereas most of the world suffered renewed setbacks.

3 Methodology

To prove our hypothesis about contagion, we first cluster the different stock markets by SOM and then we use the data from different clusters to study the contagion.

In this section, we detail our new methodological approach. We first explain the manner in which SOMs work to clarify the formation of clusters, then, revise the contagion tests following Forbes and Rigobon (2002).

3.1 Self-organizing Neural Networks

Artificial neural networks originated in the 1960s (Minsky & Papert, 1969; Rosenblatt, 1958), but were fist used in the 80s (Hopfield, 1984; Kohonen, 1982) as an alternative to the prevailing Boolean logic computation (Vesanto & Alhoniemi, 2000).

Basically, there are two kinds of neural networks: supervised and self-organising networks. The first is a universal function "approximator" (Matin & Sanz, 1997; Funahasi, 1989), used both to adjust functions and to predict results. The latter are data pattern classification networks. These kinds of networks identify similar patterns within a pool of data and group them based on these similarities (Martín & Sanz, 1997).

The first self-organising networks were so-called "Competitive Networks", which include an input and an output layer, each layer comprised of a group of cells. Model inputs are introduced through the input layer cells. Each cell in the input layer is connected to each of the cells in the output layer by means of a number, called a synaptic weight. (Willshaw & Malsburg, 1976).

[insert Here Figure 1]

The goal of the network is to find out which cell in the output layer is most similar to the data introduced in the input layer. For this purpose, the model calculates the Euclidean distance between the values of the input layer cells and the values of the synaptic weights that connect the cells in the input layer to those of the output layer. In the first step, these synaptic weights are random numbers.

The cell in the output layer that shows the least distance is the "winner" or Best-Matching Unit (BMU), and its synaptic weights are then adjusted using the learning rule, to approximate them to the data pattern in the input cells. The result is that the Best Matching

Unit has greater possibilities of winning the competition in the next submission of input data; or fewer if the vector submitted is different. To conclude, the cell has become specialised in this input pattern.

Kohonen (1982, 1989, 1990, 1997) introduced the neighbourhood function to competitive networks, creating so-called Self-Organizing Feature Maps or SOMs. Kohonen's innovation consisted of incorporating to the winning cell a neighbourhood function that defines the surrounding cells, altering the weights of both the winning cell and of other cells in the neighbourhood. The effect of introducing the neighbourhood function is that cells close to the winning cell or BMU become attuned to the input patterns that have made the BMU the winner. Outside the neighbourhood, cell weights remain unaltered. For a SOM-type self-organising network to be able to classify it must have the capacity to learn. We divide the learning process into two stages:

1. Classification, to identify winning neurons and neighbouring neurons.
2. Fine adjustment, to further specialise winning neurons.

The mechanics of self-organising maps begin by allocating random synaptic weights W_{ijk} to link the input layer and the output layer. Next an input data pattern, X(t), is introduced, and each neuron in the output layer calculates the similarity between its synaptic weight and the input vector by means of the Euclidean Distance[5] represented in Equation 1.

Equation 1

$$d = \sqrt{\sum_{k=1}^{N} \left(W_{ijk} - X_k\right)^2}$$

The output network neuron that shows the least distance to the input pattern is the winning neuron, g*. The next step is to update the weights corresponding to the winning neuron (W_{ijk}) and its neighbours, using the following equation:

Equation 2

$$W_{ijk}(t+1) = W_{ijk}(t) + \alpha(t) \cdot h(|i - g^*|, t) \cdot \left(X_k(t) - W_{ijk}(t)\right)$$

[5] There are other measurement criteria, such as the Manhattan distance or the Scalar product. However, the most commonly used is the Euclidean distance.

Where $\alpha(t)$ is a learning term that takes values comprised between 0 and 1. Where the number of iterations exceeds 500, then $\alpha(t)$ tends to 0. Equation 3 is usually used to calculate $\alpha(t)$.

Equation 3

$$\alpha(t) = \alpha_0 + (\alpha_f - \alpha_0) \cdot \frac{t}{t_\alpha}$$

Where α_0 is the initial rate, α_f the final rates, which usually takes values amounting to 0.01, t is the current situation and t_α is the maximum number of desired iterations.

The function $h(|i - g^*|, t)$ is the neighbourhood function, and its size is reduced in each iteration. The neighbourhood function depends on the distance to BMU and on the neighbour ratio. This function tells us that the neighbourhood function decreases when the distance to the winning cell increases. The further away from the winning neuron, the smaller the cell's neighbourhood function, depending also on neighbour ratio R(t), which represents the size of the current neighbourhood.

$$h(|1 - g^*|, t) = f[R(t)]$$

To calculate neighbourhood, Step functions or Mexican hat-type functions are used. The neighbour ratio R(t) decreases in time. Below is a commonly used equation that reduces the neighbour ratio in time:

Equation 4

$$R(t) = R_0 + (R_f - R_0) \cdot \frac{t}{t_R}$$

R_f is the final ratio, which takes a value equal to 1. Likewise, t_R is the number of iterations required to reach R_f.

In the fine adjustment stage, α is equal to 0.01, and the neighbour ratio is equal to 1. The number of iterations is proportional to the number of neurons and separate from the number of inputs. Usually, between 50 and 100 iterations are sufficient.

The greater the number of identical patterns, the greater the number there will be of neurons that specialise in such a pattern. The number of neurons specialised in recognising an input pattern depends on the likelihood of such a pattern. The resulting

map therefore approaches a probability density function of the sensory space. The number of neurons concentrated in a certain region show the greater likelihood of such patterns.
After declaring which is the winning neuron (*Best-Matching Unit*, BMU), the SOM's weight vectors are updated, and their topological neighbours move towards the input vector, thus reducing the distance. This adaptation generates a narrowing between the winning neuron and its topological neighbours in respect to the input vector. This is illustrated in Fig. 2, where the input vector is marked by an X. The winning neuron is marked with the acronym BMU. Observe how the winning neuron and its neighbours get closer to the input vector. This displacement is reduced to the extent that the distance from the BMU is greater.

[insert Here Figure 2]

SOM is a well-known methodology in several fields. It is used in a wide range of activities (Hertz, Krogh & Palmer, 1991). For example: Communication security (Barrera, Kayacik, Oorschot & Somayaji, 2010), Water resource problems (Kalteh, Hjorth & Berndtsson, 2008), Genetics (Hewitson & Crane, 2002), Integrated modeling of business systems (Ferstl & Sinz, 1998), Data Mining (Azcarraga, Hsieh, Pan & Setiono, 2008) (Lawrence, Almasi & Rushmeier, 1999), and Market segmentation (Bloom, 2005;Lee, Gu & Suh, 2006). Nevertheless, there are few applications in finance apart from credit scoring.
Specifically, The European Central Bank has supported a credit scoring study with SOM. Along the same lines, García-Estévez and Carballo (2015) have also used this algorithm to establish a credit scoring alternative to the current methods. One of the first applications in finance was developed by Serrano-Cinca (1996). Kourtit, Nijkamp and Arribas (2012) made a study about smart cities with SOM.
SOM networks are particularly useful for establishing unknown relations between datasets. Datasets that do not have a known preset order can be classified by a SOM network. SOM can provide an easily interpretable non-linear description of the multidimensional data distribution on a two-dimensional plane without losing sight of individual indicators. Applied to the data and the problem studied here, the input variables are the five stock market characteristics explained in the previous section, and the concept of neighborhood of a stock market represents the similarity to the current macro-financial

conditions. The output will be a two-dimension map of stock exchanges that summarizes the information on the five key variables.

First, a large set of prototypes or cells—much larger than the expected number of clusters—is formed using the SOM (Figure 3). The prototypes can be interpreted as "protoclusters," which are combined in the next step to form the actual clusters (Vesanto & Alhoniemi, 2000).

Each data vector of the original data set belongs to the same cluster as its nearest prototype.

The primary benefit of this two-level approach is the reduction of the computational cost. Another benefit is noise reduction. The prototypes are local averages of the data and, therefore, less sensitive to random variations than the original data.

[Insert Figure 3, About Here]

Sarlin and Peltonen (2013) used this methodology to analyze the economic crisis, and they found that a crisis in one position on the map indicates the propagation of financial instabilities to adjacent locations. However, in our paper, each map location indicates similarities with the closest markets.

By applying ANNs to financial markets, Cottrell, De Bondt and Grégory (1996) simulated different interest rate scenarios, and Martín and Serrano-Cinca (1994) analyzed the Spanish banking crisis. In addition, most of the research has been applied to financial asset forecasting, specifically, commodity futures, bonds, stocks or indices: Barr and Mani (1994), Donaldson and Kamstra (1996), Grudnitsky and Osburn (1993), Kryzanowski, Galler and Wright (1993), Cao, Leggio and Schnierderjans (2005), and Jasic and Wood (2004). More closely related to our work, Kolarun (2010) and Ao (2003a, 2003b) studied regional markets.

The algorithm assigns a node in the output grid to each input vector. The input vectors placed near other input vectors will be assigned to nearby nodes of the output grid. The mechanics of SOM allocates random weights to link the input layer and the output layer. An input data pattern is then introduced, and each cell in the output layer calculates the similarity between its synaptic weight and the input vector by means of the Euclidean distance. The output network cell that indicates the lowest distance to the input pattern is the BMU (Best Match Unity). Next, the updating of the weights is conducted using the neighborhood function, so that the updating decreases as the distance to the winning cell increases. The farther away from the BMU, the smaller is the cell neighborhood function. The greater the number of identical patterns, the greater the number of cells that will specialize in this pattern. Therefore, the resulting map approaches a probability density function of the sensory space. The number of cells concentrated in a certain region indicates the likelihood of these patterns.

The neighborhood function of stock markets represents their similarities in the current macro-financial conditions. Salin and Peltonen (2013) found that a crisis in one position on the map indicates the propagation of financial instabilities to adjacent locations. In our paper, a position in the map shows the similarity between adjacent markets.

3.2 Mapping and Contagion

We begin the empirical research by normalizing the five variables using a logistic normalization procedure to prevent the size of the variables from affecting the results. The scaling of the variables is of special importance because a Euclidean metric is used to measure the distances between vectors. Logistic normalization[6] ensures that all values are within the range [0 1]. The transformation is essentially linear in the middle range and has a smooth nonlinearity at both ends, which ensures that all values are within the range. Next, we draw the map based on Euclidean Distance. The Stock markets placed near other markets present a short Euclidean distance from these. The model shows the markets subdivided into four clusters or groups, from the most to the least developed markets. The markets with equal characteristics are in the same cell, and similar markets are located in the neighboring cells. The markets situated in cells that are close to the border of the group have a higher probability to change to another group than the markets that are situated in the center of the group.

[6] The data are first scaled as in variance normalization: $\hat{x} = (x - \bar{x})/\sigma_x$. This is followed by the logistic function $x' = 1/[1 + exp(-\hat{x})]$. The transformation parameters are the mean value $\bar{x}$ and standard deviation σ_x of the original values x, as in the standard normalization procedure.

Although the projection provides an approximate idea of the clusters, the U-matrix is the most widely used technique to show the distances of each map unit to each of its immediate neighbors (Vesanto, 1999). We indicate the unified matrix map based on all the variables. This map visualizes the distances between the neighboring map units or cells and helps clarify the cluster structure of the map (Vesanto, Himberg, Alhoniemi & Parhankangas, 2000). This map is unrelated to the x-y Cartesian axes, in which the vertical and horizontal axes refer to numerical values. In other words, a specific unit or cell has a meaning according to its size, not its position on the map. (Figure 4 left)

The size of each cell indicates its Euclidean distance to its closest neighbor. A larger size indicates a lower distance between cells. Each unit gathers a set of markets with equal data, whereas the markets with similar data are in the nearest neighborhood.

Figure 4 (right) indicates the overall result of our model: the market distribution according to the variables' likeness. To draw Figure 4 (right) from the information in Figure 4 (left), we use k means clustering, which is a method to group variables using centroids. The goal is to establish centroids of different categories and then classify them by proximity. The first step is choosing k centroids in a random manner; however, they must be far from one another. The second step is classifying the different cells depending on the nearest centroids. The algorithm then re-computes the position of the centroids using the cells that belong to each class; it repeats these iterations until the centroids no longer move. Finally, the Davies-Bouldin index is calculated for each clustering (Jain & Dubes, 1988; Davies & Bouldin, 1979). This index is suitable for evaluation of k-means partitioning because it gives low values, indicating good clustering results for spherical clusters. (Vesanto & Alhoniemi, 2000). The result is the right-located map in Figure 4 (right), which indicates the different groups of this model based on the 65 cells as the outcome of the specific sample size.

[Insert Figure 4 about here]

Markets located within the same cell feature the shortest Euclidean distance between their variables. As observed in Figure 4 (right), the map is divided into four zones: A, B, C, and D. Zone D gathers all the cells with the shortest Euclidean distances between them; zones B and C, the cells with intermediate distances; and zone A, the cells with the largest distances.

Next, to describe the potential contagion, we proceed to calculate an index for each of the formed clusters. The index will measure the evolution of the weighted monthly domestic capitalization.

First, to find the monthly weight for each market, x^t_j, within a specific group, we divide the market's monthly domestic capitalization, cap^t_j, by the total capitalization of its own group:

$$x_j^t = \frac{cap_j^t}{\sum_{i=1}^{n} cap_i^t} \qquad [1]$$

Because each group composition changes from one period to another, we use the clusters as if they were portfolios. Next, for every period we calculate the monthly market return within a specific group. Finally, we multiply these monthly returns by their weights to find the monthly return of each cluster or group, r_{tj}, defined by the model.

$$r_{cl}^t = \sum_{j=1}^{n} x_j^t r_j^t \qquad [2]$$

being $r_j^t = \frac{cap_j^t - cap_j^{t-1}}{cap_j^{t-1}}$.

As stated, in accordance with Forbes and Rigobon (2002), we define contagion as a strong co-movement between two groups after a shock, and we calculate the volatilities of each group.

[Insert Figure 5 about here]

The volatilities of all groups increase after the subprime crisis of 2008 (Figure 5). However, this does not mean there is a contagion effect; the volatility could just be a sort of interdependence. To determine whether there is a contagion effect, we first divide the time into two non-overlapping segments, one stable before the subprime crisis and other unstable after.

The correlation coefficients between markets will enable us to analyze the existence of contagion. The test measures the correlation between two markets during a stable and an unstable period. If the coefficient significantly increases, it can be surmised that the transmission mechanism between the markets has been strengthened; therefore, there is evidence of contagion.

After a shock, markets tend to be more volatile. The market returns' heteroscedasticity may cause an incremental bias of the correlation coefficients' values between markets after the crisis. Without adjusting the coefficients, it is impossible to know whether the increment has been due to an increase in the unconditional correlation or simply to a surge of volatility. We have used the VAR procedure to estimate the correlation coefficients. The coefficient of correlation is calculated as:

$$\rho = \frac{\rho^*}{\sqrt{1+\delta[1-(\rho^*)^2]}} \qquad [3]$$

Being: $\delta = \frac{\sigma^h_{xx}}{\sigma^l_{xx}} - 1$

Where σ^h_{xx} is the variance of market x during the high volatility period generated after the event, and σ^l_{xx} is the variance of market x during the stable period before the event. ρ^* is the conditional correlation coefficient.

We conducted different tests using 2, 3, and 4 lags to control serial correlation.

Finally, we use the *t- test* to observe whether there is a significant increment of the correlation coefficient during the most volatile period. If we call $\rho\phi$ the coefficient for the entire period, and $\rho\eta$ the coefficient for the most volatile period, the hypothesis of our test is:

H₀: $\rho\phi > \rho\eta$ There is no contagion evidence

H₁: $\rho\phi \leq \rho\eta$ There is contagion evidence

4 Results and Discussion: Stock markets' mapping, clusters' dynamics and contagion

In Figure 6, we include the data distribution maps and the descriptive statistics for each of the five variables. The five figures represent the different components linked by position. Each hexagon corresponds to the same unit in the maps. Each stock exchange in each of the four periods has been allocated to one cell according to the classification model. The grey code represents the de-normalized quantification of each variable. Thus, dark and light shades indicate low and high-level variables, respectively. Each component indicates the values of one variable in each map unit.

The resulting distribution in Figure 4 (right) is used for each map in Figure 6 to establish the level of variables in each group. Therefore, we can observe that cells/markets in group A indicate the highest values for three of the five variables: DMC, ET and LF. In addition, group B indicates the highest values for LD and NNC, whereas group D indicates the lowest values of all variables.

[Insert Figure 6 about here]

We observe that the most distinctive and influential variable in our model is DMC and that ET leads to approximately the same classification.

Next, in Table 2 we display the descriptive statistics of all variables, as well as the Euclidean distances.

[Insert Table 2 about here]

We observe that group A presents the highest levels of capitalization, turnover, and foreign companies listed, whereas group D presents the lowest levels in all variables. The Average DMC in group A is more than 32 times the average amount in group D and more than three times that in the following group B. Nevertheless, the most clear-cut differences between zones can be found in the variables ET and LF. ET for zone A is more than 90 times that of D, 14 times that of C, and nearly six times that of B. This result means a very high level of transactions in group A, which are partially due to a high LF (LF for zone A is 43 times that of D and more than 11 times that of B); this indicates a greater presence of foreign companies. Exchanges in zone C are comparatively more international than those in B, as measured by their high level of LD.

Although Group B has lower levels of capitalization and turnover, it indicates a higher level of new IPOs than Group A. As we shall comment later, markets in this group present more domestic market dynamism and comparatively less internationalization. LD is the least influential variable in the model.

The average Euclidean distance in group A is more than two times that of group D, although it is like that in groups B and C. Markets in zone D present more similar values, resulting in a greater likeness. In contrast, all the other groups indicate similar average Euclidean distances, which are higher than that of group D and, accordingly, present a higher level of disparity between exchanges.

[Insert Table 3 about here]

In Table 3, we observe the average distances between the groups; markets in group A with the highest distance have more “personality”, that is, more differences with the remainder. In Figure 7, we visualize the distribution of the 65 cells that comprise the map in a 3-D graph. The distance between the groups can also be clearly appreciated. The markets belonging to group D are the most similar, and they have the shortest Euclidean distance and more connectivity. In contrast, group A is located at the farther end of the map.

[Insert Figure 7 about here]

To gain insight into the differences and the time evolution of the five variables, in Figure 8 we graph the average values during each period. We observe that the increase in DMC has been steady during the four periods, and the main change has occurred at the ET variable, which increased dramatically for group A during the 2003-2007 period. Variables LD as well as LF have remained stable.

[Insert Figure 8 about here]

We note that the leadership of group A has been maintained during the four periods, although during the subprime crisis period, Groups B and C led in corporate activity (NNC). However, group A regained its leadership during the recovery period of 2010-2013. Group D is the smallest and most stable during the entire period analyzed, whereas groups C and B exhibit certain dynamics although the distances remain constant with group A mainly in terms of DMC, ET, and LF.

Next, we proceed to visualize each stock exchange in its corresponding cell on the map (Figure 9). The cells in Figure 8 are numbered from 1 to 65 to facilitate the analysis. To identify each cell in the map, we have indexed them, starting with the top left cell with the number 1 and moving completely down to the bottom right cell with the number 65.

To have a better understanding of Figure 9, we have separated the four groups, highlighting those stock exchanges that experienced classification changes. We use dotted arrows to define the stock markets that move between groups and a thin arrow to indicate the markets that move backwards.

[Insert Figure 9 about here]

As previously noted, the A area covers the most capitalized and liquid stock markets of the world. Only the NYSE, NASDAQ, and LSE are in this area. None of the other markets enter or exit this zone, and the distance with the other exchanges remains stable.

Group B is mainly comprised of Asian/Pacific and European markets. Except for Tokyo, which does not change groups during the entire period analyzed, the remaining markets moved to this group from C and D: India, Korea, Hong Kong, Shanghai and Shenzhen, Spain and NYSE Euronext, in addition to TMX from Canada, and an additional market from Australia.

The C group is a transit point between groups D (the least evolved markets) and B, in which all the markets move except the Deustche Borse, Singapore, and Luxembourg. Markets that pass through this area at some point include Shenzhen, Korea, Spain, Hong Kong, and NYSE Euronext. Malaysia, Taiwan, Warsaw, and the North European markets (NASDAQ) join the C area from group D in 2008 and remain there.

Finally, group D is the largest group comprised of smaller capitalization markets. The two African markets that enter our sample, Johannesburg and Egypt, remain in this group in all periods.

We observe that the four groups are very much related to stock exchange capitalization without experiencing dramatic changes during the fourteen years. Without offering an exhaustive description of the changes, we discuss certain elements related to the countries' internal and external demand, and to their financial development based on information from the Financial Development Report.[7] The Report ranks countries based on the competitiveness of their financial systems, and those in the highest positions coincide with the stock exchanges of those countries included in zones A and B.

According to this report, markets that move into groups C and B exhibit a high level of non-banking financial services (IPOs, M&A activities, insurance, and securitization) in recent years, as also observed in Figure 8. Group B (and C) presents the highest values for LD and NNC that define the dynamic markets related to internal demand. It can be observed that many of the exchanges belong to countries with large and growing internal

---

[7] The Financial Development Report is issued annually by the World Economic Forum and summarizes the countries' financial situation according to seven pillars: institutional and business environment, financial stability, banking and non-banking financial services, financial markets and financial access.

demand and trade balances, indicating a better financial condition, e.g., China, India, Malaysia, and Australia.

The Spanish market is the sole one that moves from D to B in two consecutives steps, mainly due to the IPO activity, the rapid growth of internationalization (due to certain fiscal advantages[8]) and the capitalization of certain listed companies, without being affected by the sovereign debt crises. The Polish and Nasdaq Nordic[9] markets move from group D to C. We emphasize that half of them belong to ex-communist republics whose stock exchanges date back to the XIX century and are in the process of recovering their former status. Finally, NYSE Euronext[10] first advanced from the C to the B Group, then experienced a setback during the European sovereign debt crisis period, and finally regained its position in the B group during the final 2010-2013 period.

The Asia/Pacific area is the most relevant in terms of dynamics and numbers. Two Chinese markets, Shenzhen and Shanghai, as well as Hong Kong and Korea, evolved from D to B. India, Taiwan and Malaysia joined group C from D, whereas Australia also changed from C to B.

We observe that an area covering parts of group B and the entire group A in the lower right of the map belongs to the countries with the highest rent per capita. These countries, whose economic growth has been flat in recent years, are the US, Japan, Singapore, and certain key European exchanges, such as those in the UK and Germany, with the highest levels of capitalization and internationalization. They maintain their leadership due to their developed financial markets, institutional environment, and financial stability. These senior and long-standing financial centers have recently exhibited comparatively less dynamism in corporate activity than other markets. The internal demand of these countries is smaller than that of the Asian markets, and their growth rate is below the world average. Nevertheless, these markets' leadership relies on factors such as efficiency, transparency, reliability, low margins, liquidity, and reputation.

---

[8] Spanish tax law affecting goodwill since 1st January 2002.

[9] Nasdaq OMX Nordic operates the Helsinki, Copenhagen, Stockholm, Iceland, Riga, Tallinn and Vilnius stock exchanges.

[10] Euronext is the first pan-European exchange, spanning Belgium, France, the Netherlands, Portugal and the UK.

Therefore, the Asia/Pacific stock exchanges have developed and grown more than the European or American exchanges during the last fourteen years. At the same time, although group A has not moved during the entire period, there have been changes in the other groups, which indicates an increase in dynamism, particularly in groups B and C. Some exchanges will likely join group A in the future and challenge the actual leaders.

In Figure 10, we observe the connectivity and inner relations between nodes based on Euclidean distances, observing that Group D on the left exhibits the highest level of connectivity. This finding strengthens the idea that the stock markets are more like each other; thus, the higher the number of links, the stronger their likeness.[11]

[Insert Figure 10 about here]

In accordance with the method noted in the previous section, we identify a key macroeconomic event, the subprime crisis. We observe (refer to Figure 5) that the volatilities increase for the four groups immediately after the 2008 crisis. The goal is to determine whether this common surge is due to a contagion effect or to a high level of market co-movement (called interdependence by Forbes and Rigobon, 2002). To determine this, we divide the period into two non-overlapping segments, one stable before the subprime crisis and the other unstable after.

We calculate the unconditional correlation coefficients for each period. If the coefficient $\rho\eta$ (the coefficient for the most volatile period, after the shock) is higher than $\rho\phi$ (entire period coefficient), then there is contagion evidence, and therefore we accept H_1 and reject H_0.

Given that the subprime crisis appeared in the markets located in group A and affected the banks and financial institutions trading in these markets (US and UK), we will assess the potential contagion effects from A to B, C and D (see Table 4).

The coefficients for the stable, volatile, and total periods are shown in Table 5. The critical value of t at a 5% level is 1.65, and any value greater than the critical value denotes contagion, (C) whereas any lower value denotes no contagion (N).

[11] We have conducted a robustness test, available upon request, which shows similar results after adding and changing variables, and the number of periods.

[Here Table 4]

[Here Table 5]

It can be observed that, independent of the lag used, results remain stable: Group A transmits the crisis to group B and to group D (less developed and more dependent markets). However, there is no direct contagion from A to C. We recall that group B is composed of the main European and Asian exchanges during the 2008-2013 period.

Then, we observe that B affects group C (other European and Asian markets) and D again, and finally C passes on the contagion to D. These movements reinforce our previous results about market classification obtained using SOM methodology. Group B is the closest group to the stock markets in A, and it is the first to experience contagion.

In addition, as the average high level of correlation coefficients for the total period denotes, stock exchanges around the world have great interdependence; then, it is reasonable to believe that the contagion affects groups of exchanges and not merely stock exchanges one by one.

Because the contagion results are critically dependent on the way the classification into groups has been conducted, and with the intention to add robustness to our results, we relate our work to two other classifiers. There are many financial institutions that classify world markets and stock exchanges; however, they barely agree on their results. For instance, a market can be defined as emergent by one classifier and frontier (i.e., less developed but with potential) by another classifier. Given the lack of a uniform procedure and to avoid confusion, we will compare our results with those of two very different studies by Morgan Stanley Capital International (MSCI) and by Mantegna (1999). MSCI, a key classifier in the international financial arena, follows four criteria: economic development, size, liquidity, and accessibility. Although the division is arbitrary, it is used in the financial industry, and it is easy to understand. MSCI separates the markets into three distinct groups: developed, emergent, and frontier.

The other classifier, developed by Mantegna, uses the correlation matrices of 24 stock exchange indices from 20 countries from Europe, Asia, and America within the framework of the random matrix theory. Mantegna defines a subdominant ultrametric space to find a matrix distance to determine the Minimum Spanning Tree (MST) that connects stock exchange indices. In accordance with the graph theory, Mantegna obtains a hierarchical organization of indices.

The criteria that MSCI, Mantegna, and our work follow are dissimilar; Mantegna uses indices and MSCI markets / countries and certain economic variables, whereas we made use of market and micro market variables. However, we observe a certain common ground. The Mantegna classification finds regional and stable links between different exchanges. The links indicate four distinct clusters: Japan (Topix and Nikkei), the USA and Canada, Asian Pacific, and Europe, omitting certain markets from the clusters, namely, Mexico, Chile, and India. Conversely, the MSCI market taxonomy places all developed markets in one group and divides the remainder into subgroups: emergent and frontier.

In contrast to the other models, our model does not use indices or macro data; instead, it uses five variables. Our taxonomic output, as previously noted, has a transversal and more dynamic character, grouping certain regions (Asia, Europe) while at the same time placing together certain emergent, frontier, and developed, albeit small, stock exchanges (refer to Oslo, Cairo, and Mexico in group D).

Group A is comprised of three exchanges, which are classified as developed by the MSCI. Groups B and C are the Asian and European, mainly developed and emergent exchanges; these include giants such as Tokyo (B) and Singapore (C). Group D is a hodgepodge of exchanges from all around the globe, comprised of large emergent and frontier markets, including a few small European markets. This large group is not considered by Mantegna; that is, it is designated as an off-the-clusters market, whereas the MSCI classifier solely places frontier exchanges (developing with potential). However, ours includes certain small developed exchanges. Nevertheless, MSCI considers Argentina the sole market to include in group D, whereas in our study, not surprisingly due to its poor performance, it remains in that group during the entire 14-year period.

5 Conclusions

The use of SOM allows us to create an innovative map of stock exchanges and to study the dynamics of contagion from a non-traditional perspective: the clusters' perspective. Stock markets are leading indicators of the state of an economy, and countries with more developed financial markets have been more exposed to the current crisis. We can obtain more information regarding the dynamics of all markets when the unit of analysis is a group of exchanges with similar features instead of individual markets. SOM compiles the databases into a 2-D map, which is visual and intuitive. The map also allows us to follow the evolution of the different stock exchanges. The classification is conducted using patterns of likeness; thus, the map allows us to identify similarities and differences between stock exchanges.

Our model is the first to accomplish a dynamic classification of stock exchanges that covers a considerable number of variables and markets worldwide, regardless of whether these markets are developed, emergent, or frontier. Consequently, we found four distinct clusters defined in terms of capitalization, liquidity, and internationalization, which are not necessarily linked to different levels of efficiency or other special qualifications.

Group A is sparsely populated, with only three exchanges: London SE, Nasdaq OMX, and NYSE-Euronext. However, there are few similarities between them. This is because, in addition to their high correlations due mainly to globalization, these exchanges are very independent from each other as a result of their idiosyncrasies, i.e., their culture, rules, legislation, and micro market structure. In addition, no other market has moved in or out of the group during the 13-year period analyzed.

Groups B and C present a higher level of connectivity and similarity although they belong to the Asia/Pacific and European areas. Group D is densely populated, and although these markets are from different geographical areas, sizes, and levels of efficiency, they present many similarities in relation to the analyzed variables.

We have also compared our results to those of other classifiers. We observe that our work cut across geographic and development features, dynamically grouping markets or placing them apart.

We study the contagion between the exchanges' clusters. A key macroeconomic event is identified: the subprime crisis. We find contagion from the more developed markets in which the crisis sprouted to the other groups. Group A transmits the crisis to group B, Asia's and Europe's most developed markets, and group B affects group C, other European and Asian markets. All influence D, the less developed and more dependent markets. In addition, this paper's focus on the groups' contagion allows us to observe that stock exchanges have great interdependence among them, every time.

We find with some satisfaction that the high connectivity achieved during the last ten years, due to different causes, reinforces our group approach to the crisis spread. Additionally, this finding allows us to propose the use of this perspective for the analysis of future macroeconomic events.

6 References

Alexander, G. J., Eun, C. S., & Janakiramanan, S. (1988). International listings and stock returns: Some empirical evidence. *Journal of Financial and Quantitative Analysis*, 23(2), 135-151.

Ao, S. I., (2003a). Analysis of the interaction of Asian Pacific indices and forecasting opening prices by hybrid VAR and neural network procedure. *Proceedings of the International Conference on Computational Intelligence for Modelling Control and Automation 2003(CIMCA2003), Vienna, Austria.*

Ao, S. I., (2003b). Incorporating correlated markets' prices into stock modeling with neural network. *Proceedings IASTED International Conference on Modelling and Simulation 2003, Palm Springs, USA*, pp. 353-358.

Azcarraga, A., Hsieh, M. H., Pan, S. L., & Setiono, R. (2008). Improved SOM labeling methodology for data mining applications. In Soft Computing for Knowledge Discovery and Data Mining (pp. 45-75). Springer, Boston, MA.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129-152.

Barr, D.S., & Mani, G., (1994). Using Neural Nets to Manage Investments, *AI Expert*, February, pp. 16-21.

Barrera, D., Kayacik, H. G., van Oorschot, P. C., & Somayaji, A. (2010). A methodology for empirical analysis of permission-based security models and its application to android. In Proceedings of the 17th ACM conference on Computer and communications security (pp. 73-84). ACM.

Basu, J. K., Bhattacharya, D., and Kim, T. H. (2010). Use of artificial neural network in pattern recognition. *International Journal of Software Engineering and its Applications*, 4(2).

Bloom, J. Z. (2005). Market segmentation: A neural network application. *Annals of Tourism Research*, 32(1), 93-111.

Cao, Q., Leggio K.B. & Schniederjans M.J., (2005). A Comparison Between Fama and French's Model and Artificial Neural Networks in Predicting the Chinese Stock Market. *Computers and Operations Research*, 32: pp. 2499–512.

Cottrell, M. De Bodt, E., & Grégoire, P., (1996). Simulating Interest Rate Structure Evolution on a Long-Term Horizon: A Kohonen Map Application, *Proceedings of Neural Networks in The Capital Markets*, Californian Institute of Technology, World Scientific Ed., Pasadena.

Davies, D.L., & Bouldin, D.W., (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224-227.

De la Torre, A., Gozzi, J. C., & Schmukler, S. L. (2007). Stock market development under globalization: Whither the gains from reforms? *Journal of Banking and Finance*, 31(6), 1731-1754.

Ding, L., & Hiltrop, J. (2010). The electronic trading systems and bid-ask spreads in the foreign exchange market. *Journal of International Financial Markets, Institutions and Money,* 20(4), 323-345.

Donaldson, R. G., & Kamstra, M., (1996). Forecast combining with neural networks. *Journal of Forecasting,* 15(1), pp. 49-61.

Ferstl, O. K., & Sinz, E. J. (1998). Modeling of business systems using SOM. In Handbook on architectures of information systems (pp. 347-367). Springer, Berlin, Heidelberg.

Forbes, K. J. & Rigobon, R., (2002). No contagion, only interdependence: measuring stock market co-movements. *The Journal of Finance*, 57(5), pp. 2223-2261.

Frankel, J., & Saravelos, G., (2012). Can leading indicators assess country vulnerability? Evidence from the 2008–09 global financial crisis. *Journal of International Economics*, 87(2), pp. 216-231.

García-Estévez, P., & Carballo, A. (2015). Qualitative judgement in public credit ratings: A proposed supporting approach using Self-Organising Maps (SOMs). *Cuadernos de Economía*, 38(108), 181-190.

Grudnitski, G., & Osburn, L., (1993). Forecasting S&P and gold futures prices: An application of neural networks. *Journal of Futures Markets*, 13(6), pp. 631-643.

Hagan, M. T., Demuth, H. B., & Jesús, O. D. (2002). An introduction to the use of neural networks in control systems. *International Journal of Robust and Nonlinear Control*, 12(11), 959-985.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). Introduction to the Theory of Neural Computation, volume I of Lecture Notes of the Santa Fe Institute Studies in the Science of Complexity.

Hewitson, B. C., & Crane, R. G. (2002). Self-organizing maps: applications to synoptic climatology. *Climate Research*, 22(1), 13-26.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the national academy of sciences, 81(10), 3088-3092.

Jain, A.K. & Dubes, R.C., (1988). *Algorithms for Clustering Data*, Prentice Hall, pp. 96-101.

Jasic, T. & Wood D., (2004). The Profitability of Daily Stock Market Indices Trades Based on Neural Network Predictions: Case Study for the SandP 500, the DAX, the TOPIX and the FTSE in the Period 1965–1999, *Applied Financial Economics*, 14(4): pp. 285–97.

Jegede, O., Awodele, O., Ajayi, A., & Ndong, M. (2007). Development of a microcontroller based robotic arm. In Proceedings of the 2007 CSITEd Conference" (Information Technology).

Kalteh, A. M., Hjorth, P., & Berndtsson, R. (2008). Review of the self-organizing map (SOM) approach in water resources: Analysis, modelling and application. *Environmental Modelling and Software*, 23(7), 835-845.

Kohonen, T., (1982). Self-Organized formation of topologically correct feature maps. *Biological Cybernetics* (43), pp. 59-69.

Kohonen, T., (1989). *Self-Organization and Associative Memory*. Berlín: Springer-Verlag.

Kohonen, T., (1990). The Self-organizing Map. *Proceeding of the IEEE*, 9(78), pp. 1464-1480.

Kohonen, T., (1997). *Introduction to SOM Toolbox*. Helsinki: EET DST.

Kolarun, J., (2010). Multivariate Comparative Analysis of Stock Exchanges: The European Perspective, in Fink et al, (Eds) *Advances in Data Analysis, Data Handling and Business Intelligence. Studies in Classification, Data Analysis, and Knowledge Organization*, (pp. 461-469), Springer-Verlag Berlin Heidelberg.

Kourtit, K., Nijkamp, P., & Arribas, D. (2012). Smart cities in perspective–a comparative European study by means of self-organizing maps. *Innovation: The European Journal of Social Science Research*, 25(2), 229-246.

Kryzanowski, L., Galler, M., & Wright, D. W., (1993). Using artificial neural networks to pick stocks. *Financial Analysts Journal*, 49(4), pp. 21-27.

Lawrence, R. D., Almasi, G. S., & Rushmeier, H. E. (1999). A scalable parallel algorithm for self-organizing maps with applications to sparse data mining problems. *Data Mining and Knowledge Discovery*, 3(2), 171-195.

Lee, S. C., Gu, J. C., & Suh, Y. H. (2006). A comparative analysis of clustering methodology and application for market segmentation: K-means, SOM and a two-level SOM. In International Symposium on Methodologies for Intelligent Systems (pp. 435-444). Springer, Berlin, Heidelberg.

Levine, R., & Schmukler, S. L. (2006). Internationalization and stock market liquidity. *Review of Finance*, 10(1), 153-187.

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

Martín, B., & Sanz, A., (1997). *Redes Neuronales y Sistemas Borrosos*. Madrid: Ra-MA.

Martín, B., & Serrano-Cinca, C., (1994). Self-organizing neural networks: The financial state of spanish companies. In Refenes, A. P. (Eds) *Neural networks in the capital markets* (pp. 341-357). Wiley London.

McConnell, J. M.; Dybevik, H. J.; Haushalter, D. & Lie, E., (1996). A Survey of Evidence on Domestic and International Stock Exchange Listings with Implications for Markets and Managers Purdue CIBER Working Papers. Paper 108. http://docs.lib.purdue.edu/ciberwp/108.

Minsky, M., & Papert, S. (1969). Perceptrons: A introduction to Computational Geometry. MIT press.

Pagano, M., Randl, O., Röell, A. A., & Zechner, J. (2001). What makes stock exchanges succeed? Evidence from cross-listing decisions. *European Economic Review*, 45(4-6), 770-782.

Quinn, D., & Voth, H. J., (2008). A century of global equity market correlations. *American Economic Review: Papers and Proceedings*, 98(2), pp. 535-540.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386.

Sarlin, P., & Peltonen, T. A., (2013). Mapping the state of financial stability. *Journal of International Financial Markets, Institutions and Money,* 26, pp. 46-76.

Serrano-Cinca, C. (1996). Self-organizing neural networks for financial diagnosis. Decision Support Systems, 17(3), 227-238.

Vesanto, J., & Alhoniemil, E., (2000). Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11(3), pp. 586-600.

Vesanto, J., (1999). SOM-Based Data Visualization Methods. *Intelligent Data Analysis*. 3(2), pp. 111-126

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J., (2000). *SOM Toolbox for Matlab 5*. Helsinky University of Technology.

Willshaw, D. J., & Von Der Malsburg, C. (1976). How patterned neural connections can be set up by self-organization. Proceedings of the Royal Society B, London 194(1117), 431-445.

Table 1: Stock market sample

List of the stock markets included in the sample

Athens Exchange	Indonesia SE	New Zealand Exchange
Australian SE	Irish SE	NYSE Euronext (Euro)
Bermuda SE	Japan Exchange Group Osaka	NYSE Euronext (US)
BMandFBOVESPA	Japan Exchange Group Tokyo	Oslo Börs
BME Spanish Exchanges	Johannesburg SE	Philippine SE
Borsa Istanbul	Korea Exchange	Santiago SE
BSE India	Lima SE	Shanghai SE
Budapest SE	Ljubljana SE	Shenzhen SE
Buenos Aires SE	London SE	Singapore Exchange
Bursa Malaysia	Luxembourg SE	SIX Swiss Exchange
Colombia SE	Malta SE	Taiwan SE Corp
Colombo SE	Mauritius SE	Tehran SE
Cyprus SE	Mexican Exchange	The Stock Exchange of Thailand
Deutsche Börse	NASDAQOMX	TMX group
Egyptian Exchange	NASDAQ OMX Nordic Copenhagen	Warsaw SE
Hong Kong Exchanges	National Stock Exchange India	Wiener Börse

Table 2: Main statistics of each group and Euclidean distances within the four groups for the total period considered

The first column with letters A-D represents the groups. Average stands for the arithmetic average, STD for the standard deviation. DMC, ET, LD, LF, NNC represent the variables' domestic market capitalization, electronic trading, domestic listing, foreign listing, and number of newly listed companies through an IPO. Euclidean is the distance between the exchanges.

		DMC	ET	LD	LF	NNC	Euclidean
	Average	181,889	9,487	265	10	0	0.089
D	STD	259,174	18,148	242	15	1	0.051
	Max	1,271,308	92,750	1,151	70	3	0.212
	Min	1,259	1	10	0	0	0.004
	Average	724,172	60,732	667	107	2	0.229
C	STD	570,220	63,310	391	99	2	0.095
	Max	2,755,989	240,444	1,580	310	10	0.443
	Min	29,385	8	27	0	0	0.070
	Average	1,833,083	149,718	2,406	37	4	0.225
B	STD	926,729	110,406	1,309	43	5	0.084
	Max	3,719,542	372,790	5,106	145	16	0.428
	Min	548,490	12,941	864	0	0	0.069
	Average	5,843,447	858,665	2,241	431	3	0.211
A	STD	4,535,781	651,424	394	92	3	0.076
	Max	13,457,364	1,882,157	2,884	580	7	0.372
	Min	1,441,160	28,036	1,730	291	0	0.111

Table 3: Euclidean distances between the four groups

A, B, C, and D represent the groups. Average stands for the arithmetic average and STD for the standard deviation.

.

		B	C	D
A	Average	0.463	0.559	0.768
	STD	0.113	0.135	0.112
B	Average		0.374	0.501
	STD		0.120	0.107
C	Average			0.284
	STD			0.097

Table 4 Coefficient of correlations

We calculated an index from each cluster, which is representative of all markets which belong to it. Next, we calculated the monthly returns of the index of each cluster. This table shows the correlation coefficients of the returns between clusters. Ccorr low represents the correlation coefficient during the low volatility period; Ccorr high represents the correlation coefficient during the high volatility period and Ccorr full represents the correlation coefficient during total sample period. A, B, C and D are the clusters obtained by SOM methodology

	A	**B**	**C**	**D**
Ccorr low				
A	1.000	0.592	0.823	0.736
B	0.592	1.000	0.571	0.697
C	0.823	0.571	1.000	0.805
D	0.736	0.697	0.805	1.000
Ccorr high				
A	1.000	0.767	0.799	0.816
B	0.767	1.000	0.943	0.956
C	0.799	0.943	1.000	0.958
D	0.816	0.956	0.958	1.000
Ccorr full				
A	1.000	0.677	0.805	0.768
B	0.677	1.000	0.741	0.810
C	0.805	0.741	1.000	0.871
D	0.768	0.810	0.871	1.000

Table 5. Analysis of Contagion.

We calculate the contagion using the correlation coefficients. Firtly, we calculate the monthly returns of clusters and the correlation coefficients, ρ^{period}_{xy}, for the high volatility period, the low volatility period, and the total sample period. Certain adjustments must be made based on the VAR methodology. We conducted different tests using 2, 3, and 4 lags to control serial correlation. In each lag, we calculate the σ_{period} using VAR, in the high volatility, the low volatility and the total sample period. We calculate the conditional coefficient, ρ^{period}_{u}, in the high volatility period, the low volatility period, and the total sample period: $\rho_u^p = \frac{\rho_{xy}^p}{\sqrt{1+\left(\frac{\sigma_{high}}{\sigma_{low}}-1\right)\times\left(1-\rho_{xy}^p\right)^2}}$

Finally, we use the t- test to observe whether there is a significant increment of the correlation coefficient during the most volatile period. $t = \frac{\rho_u^{high}-\rho_u^{whole}}{\sqrt{\frac{\sigma^2_{whole}}{Nobs_{whole}}+\frac{\sigma^2_{high}}{Nobs_{high}}}}$ If we call $\rho\phi$ the coefficient for the entire period, and $\rho\eta$ the coefficient for the most volatile period, the hypothesis of our test is: *H0:* $\rho\phi > \rho\eta$ There is no contagion evidence; *H1:* $\rho\phi \leq \rho\eta$ There is contagion evidence

Clusters	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
Lags	Lag 2				Lag3				Lag 4			
STD total period	0.041	0.054	0.059	0.054	0.040	0.053	0.057	0.053	0.040	0.052	0.056	0.053
STD low period	0.035	0.049	0.054	0.049	0.035	0.045	0.052	0.049	0.034	0.043	0.052	0.048
STD High period	0.051	0.062	0.066	0.063	0.050	0.059	0.062	0.060	0.048	0.054	0.057	0.055
STD High/STD Low -1	0.441	0.283	0.219	0.274	0.419	0.322	0.195	0.227	0.392	0.238	0.112	0.130
Nº obser	164	164	164	164	164	164	164	164	164	164	164	164
Nº obser hig	65	65	65	65	65	65	65	65	65	65	65	65
		shock in				**shock in**				**shock in**		
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
ρ^{low}_{u}												
A	1.000	0.544	0.796	0.693	1.000	0.538	0.798	0.700	1.000	0.551	0.809	0.715
B	0.522	1.000	0.533	0.653	0.525	1.000	0.537	0.660	0.528	1.000	0.551	0.675
C	0.770	0.524	1.000	0.769	0.773	0.518	1.000	0.775	0.776	0.531	1.000	0.787
D	0.671	0.651	0.776	1.000	0.674	0.646	0.779	1.000	0.677	0.658	0.790	1.000
$\rho^{high}_{u} = \rho\eta$												
A	1.000	0.726	0.769	0.781	1.000	0.721	0.772	0.787	1.000	0.732	0.784	0.799
B	0.706	1.000	0.932	0.944	0.709	1.000	0.933	0.946	0.712	1.000	0.938	0.950
C	0.742	0.929	1.000	0.947	0.745	0.927	1.000	0.949	0.748	0.931	1.000	0.953
D	0.762	0.944	0.949	1.000	0.764	0.943	0.950	1.000	0.767	0.946	0.954	1.000
$\rho^{whole}_{u} = \rho\phi$												
A	1.000	0.631	0.776	0.728	1.000	0.625	0.779	0.734	1.000	0.637	0.790	0.748
B	0.608	1.000	0.707	0.774	0.611	1.000	0.710	0.780	0.615	1.000	0.723	0.792
C	0.749	0.697	1.000	0.844	0.752	0.692	1.000	0.848	0.755	0.704	1.000	0.858
D	0.706	0.773	0.849	1.000	0.709	0.768	0.851	1.000	0.712	0.778	0.860	1.000
$\sqrt{\frac{\sigma^2_{whole}}{Nobs_{whole}}+\frac{\sigma^2_{high}}{Nobs_{high}}}$	0.007	0.009	0.009	0.009	0.007	0.008	0.009	0.008	0.007	0.008	0.008	0.008
t statistic												
A	0.000	10.844	-0.697	6.010	0.000	11.383	-0.727	6.174	0.000	12.159	-0.752	6.409
B	13.855	0.000	24.031	19.273	14.024	0.000	25.006	19.611	14.481	0.000	25.668	19.942
C	-0.987	26.220	0.000	11.719	-0.996	27.777	0.000	11.884	-1.025	29.080	0.000	11.995
D	7.857	19.392	10.705	0.000	7.934	20.612	11.095	0.000	8.167	21.424	11.227	0.000
Ref	1.654	1.654	1.654	1.654	1.654	1.654	1.654	1.654	1.654	1.654	1.654	1.654
A	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
B	**C**	**N**	**C**	**C**	**C**	**N**	**C**	**C**	**C**	**N**	**C**	**C**
C	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**	**N**	**C**
D	**C**	**C**	**C**	**N**	**C**	**C**	**C**	**N**	**C**	**C**	**C**	**N**

Mapping of Stock Exchanges: Contagion from a new perspective. FIGURES

Figure 1 is a 1-column figure

Figure 1: Representation of the Self-Organizing Map (SOM).

SOM methodology has two layers. The first one is the Input Layer, in which we find the input variables of each company. Every variable is linked with every cell of the Output Layer by synaptic weights. At the beginning of the process these are random numbers. We define Best Matching Unit as the cell which presents the lowest Euclidean Distance between their synaptic weights and values of input variables.

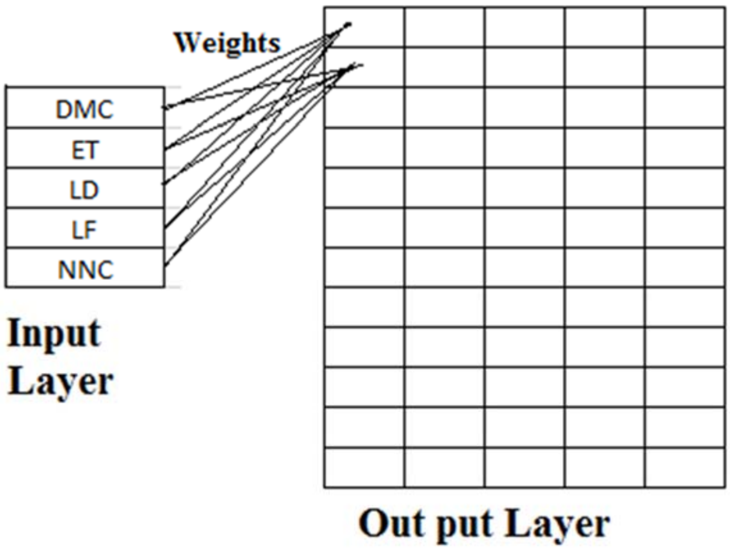


Figure 2 is a 1-column figure

Figure 2.
Updating of the winning neuron (BMU) and its neighbours, moving them towards the input vector, marked by an X. Continuous lines and dotted lines respectively represent the situation before and after the update. Source: Kohonen, 1982.

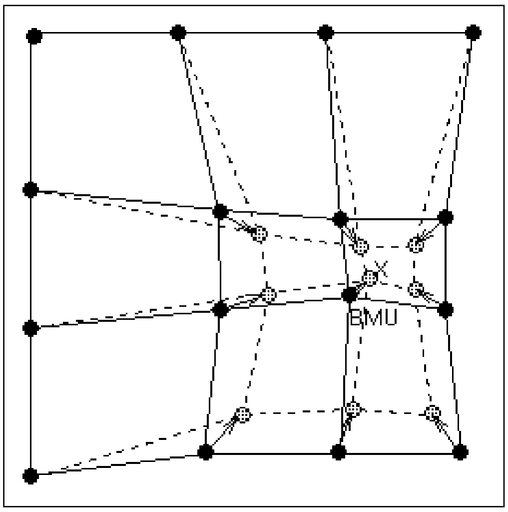


Figure 3 is a 2-column figure

Figure 3.

Based on 146 samples, we first make a set of prototype vectors using SOM (cluster level 1), followed by a clustering procedure that generates cluster level 2.
The left-chart represents all the data from the sample. The SOM methodology has placed them in different clusters. The clusters are represented by hexagons (level 1). Companies clustered in the same hexagon show similar input variables. Using k-means methodology, the model identifies the different market groups.

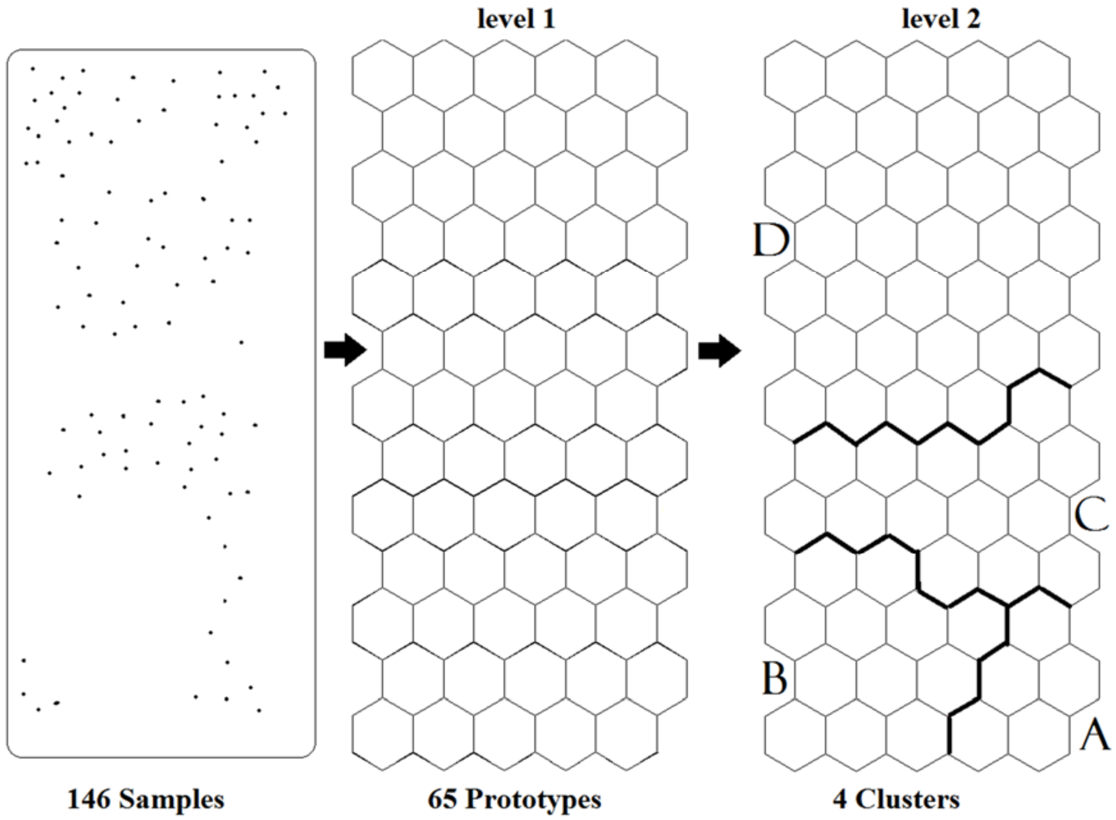


Figure 4 is a 2-column figure

Figure 4:

Unified matrix of the model (left), and cells divided into the different groups (right). The left-chart shows the number of markets in every hexagon by its size. The bigger the size, the higher the number of companies. If we compare it with the right-chart we can see that group D shows the highest number of companies, and group A the lowest number of companies

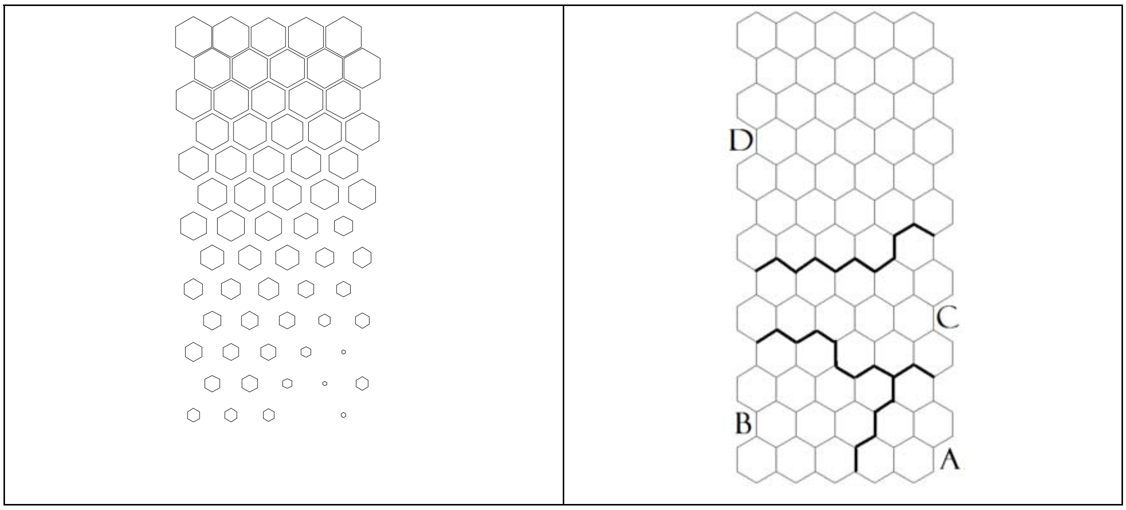


Figure 5 is a 2-column figure

Figure 5: Volatility of different groups

We build different indexes of each group, as if the groups acted as portfolios. These indexes allow us to measure the volatility. As we can see, there is one period (2009-2010) which the volatility increases in all groups.

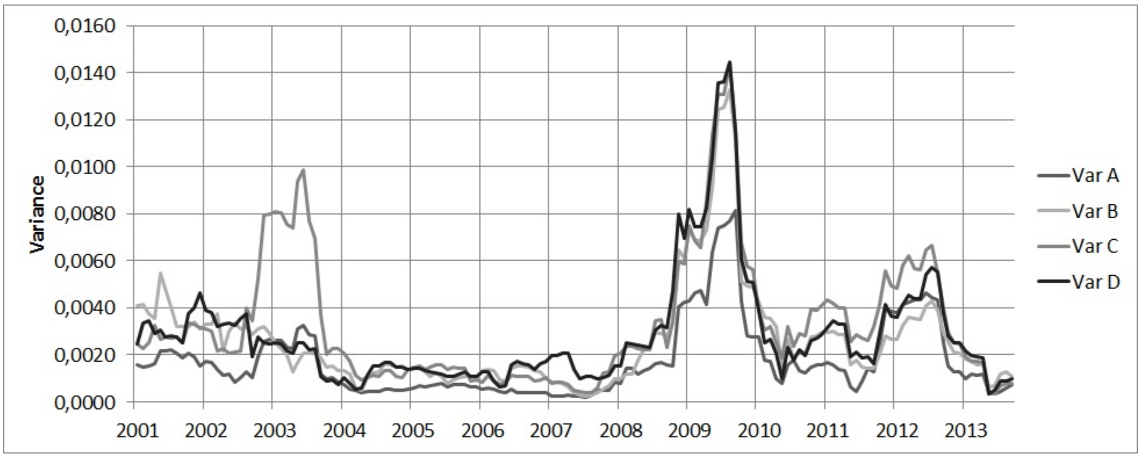


Figure 6 is a 2-column figure

Figure 6:

Classification by different variables

DMC, ET, LD, LF, NNC stand for domestic market capitalization, electronic trading, domestic listing, foreign listing, number of newly listed companies through an IPO. Each map represents the distribution of one input-variable. The clearest the cell, the higher value of variable. If a company is placed in one hexagon, this company will be placed in the same hexagon in all maps. That way, we can assess the characteristics of every company.

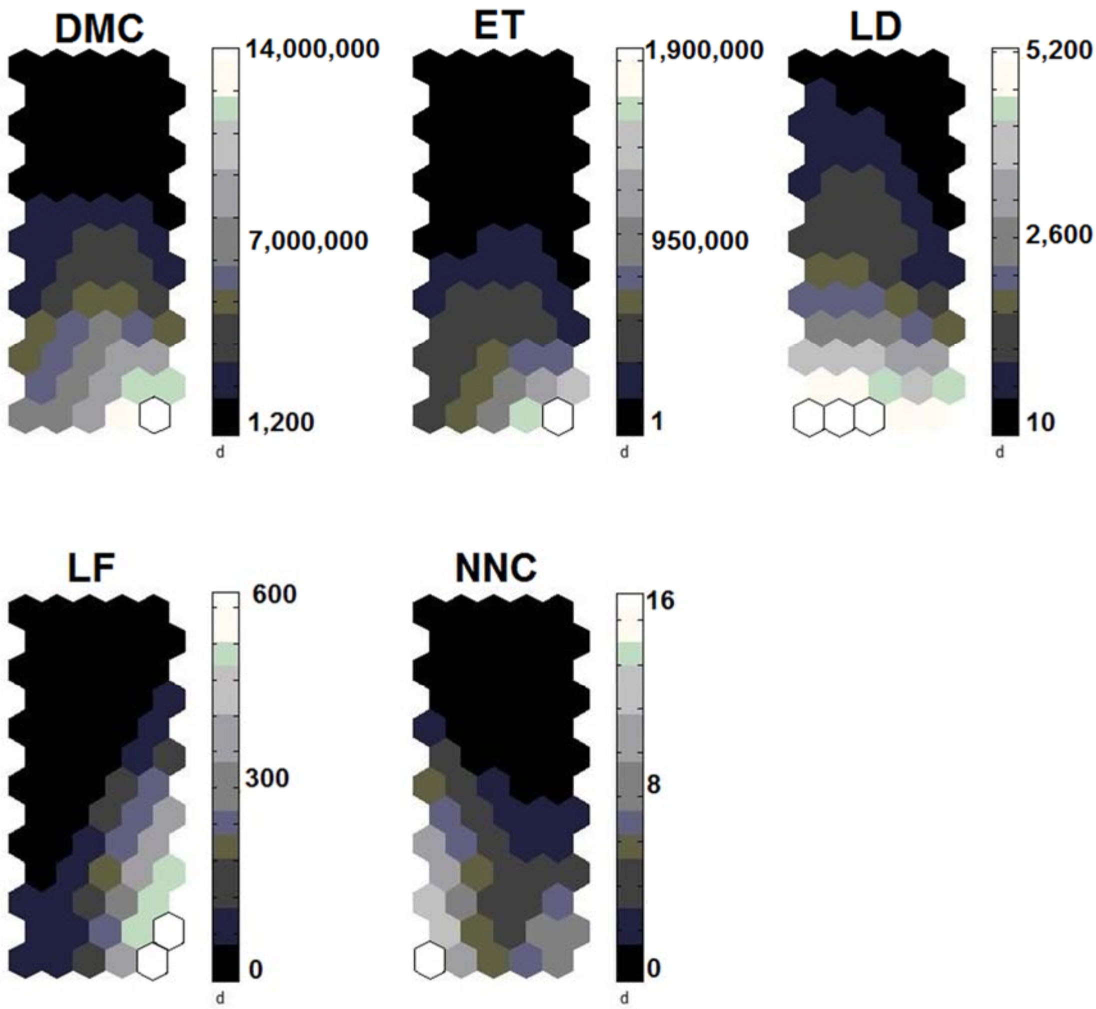


Figure 7 is a single column figure

Figure 7:

3-D distribution of the groups

Axes represent normalized Euclidean distances. We can show the total sample distributed in a 3-D chart normalizing the Euclidean distance in three axes. That way, we can observe that group A is far away from the rest of groups, which are very close, each other's.

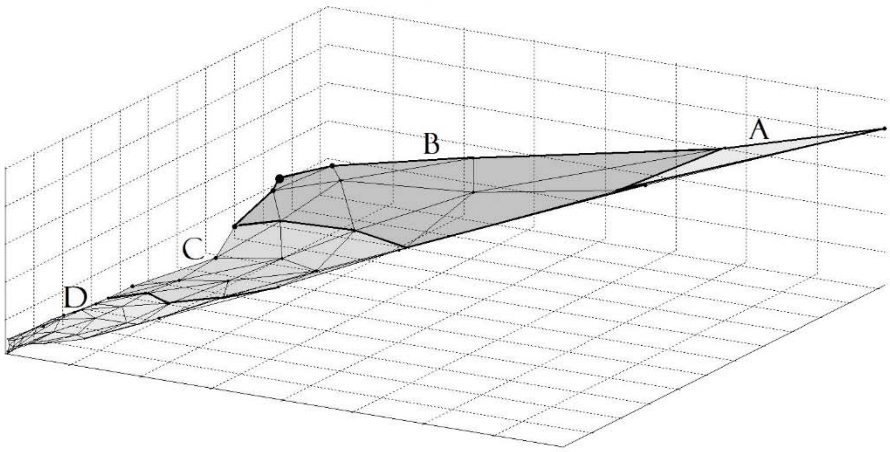


Figure 8 is a 2-Column figure

Figure 8:

Comparison of normalized variables in the four periods

DMC, ET, LD, LF, NNC stand for domestic market capitalization, electronic trading, domestic listing, foreign listing, and number of newly listed companies through an IPO. These spider-charts show the evolution of different groups in the different periods of study

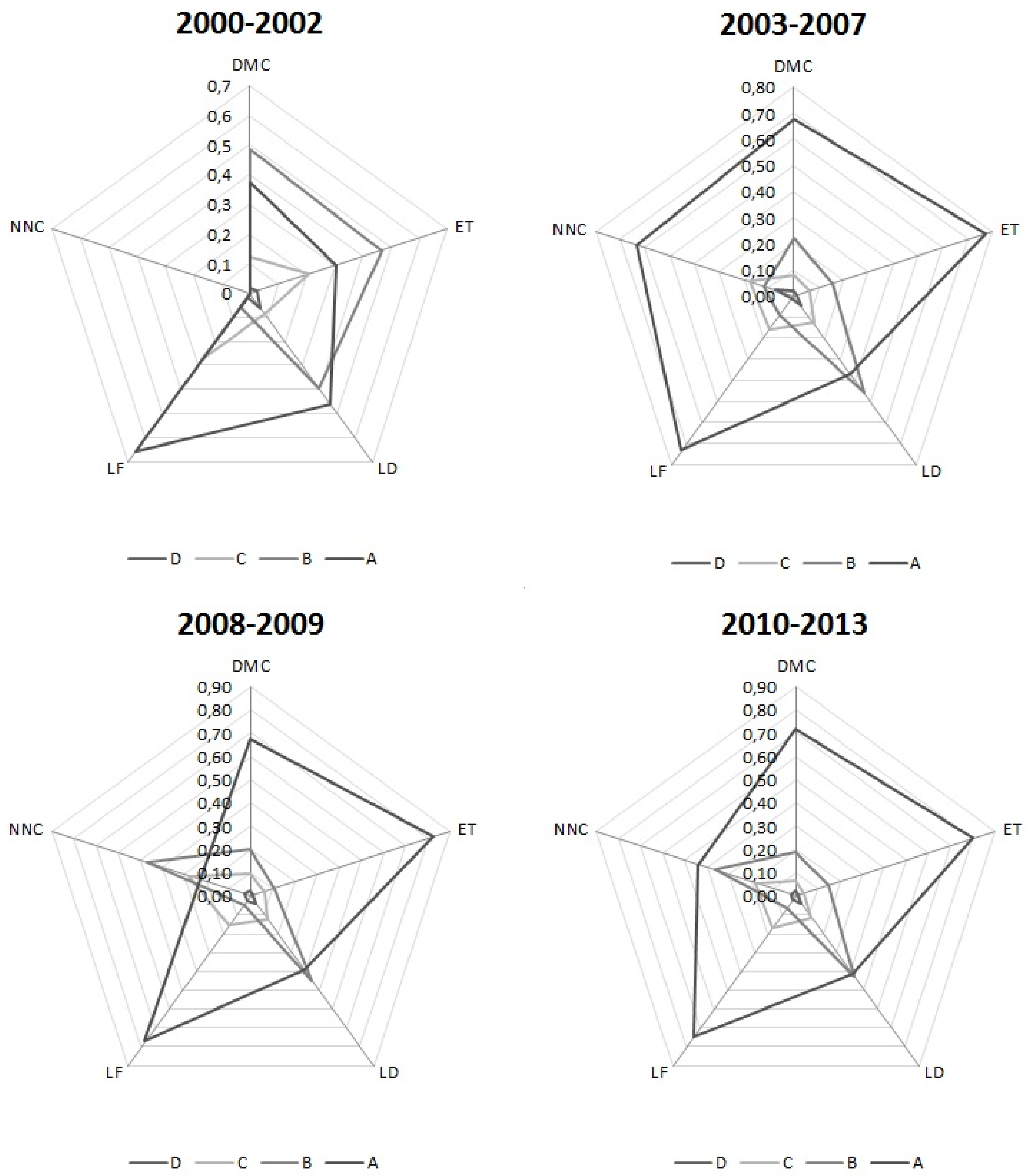


Figure 9 is a 2-Column figure.

Figure 9: Stock market mapping and evolution from 2000 to 2013. The figure represents the dynamics of the three groups. We have identified the markets that have moved from one group to another across the years. It is interesting to see that no market has jumped to group A and, except the Swiss Stock Exchange and the NYSE-euro, not one single market has gone down to the lower ranks

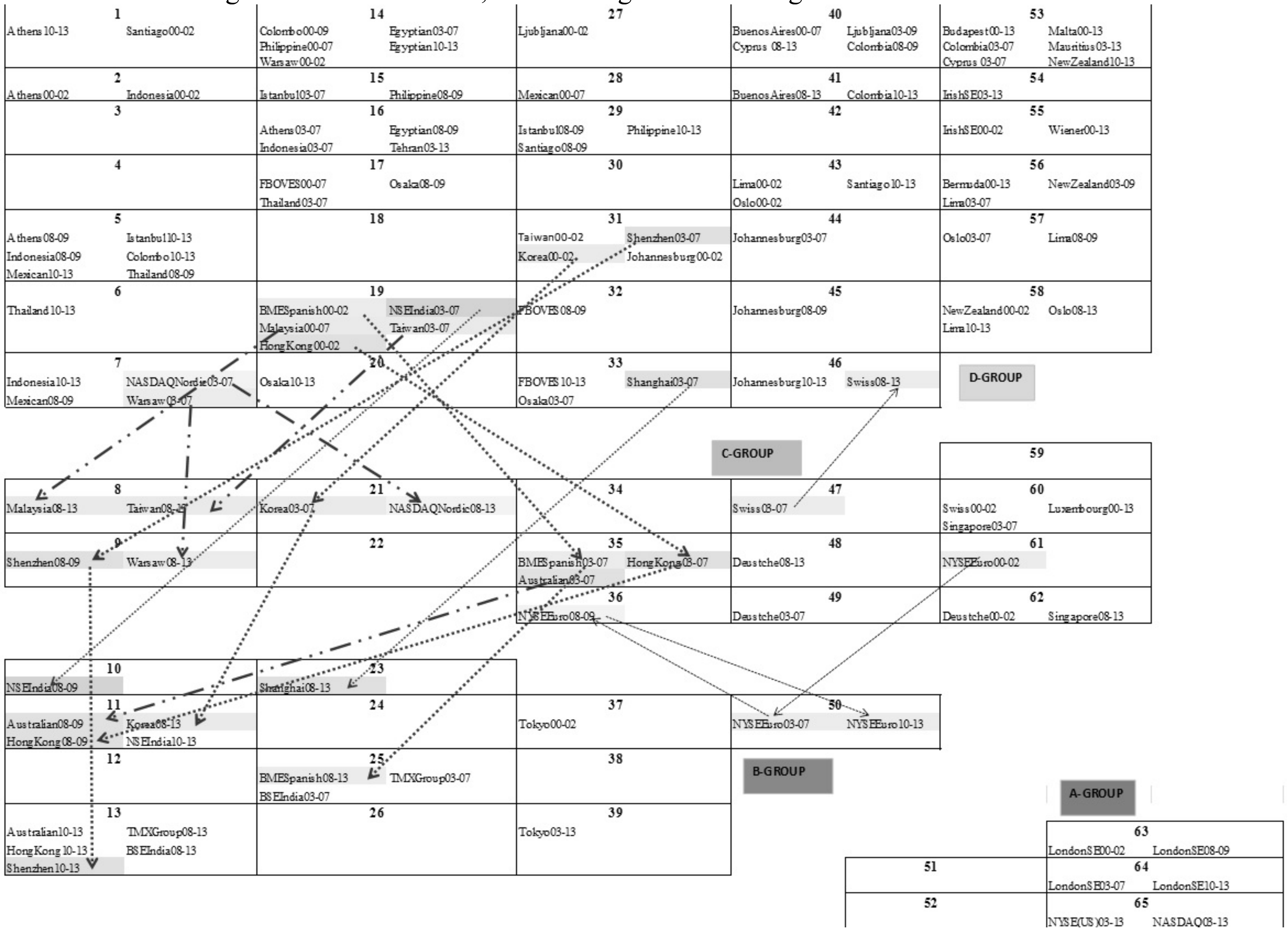
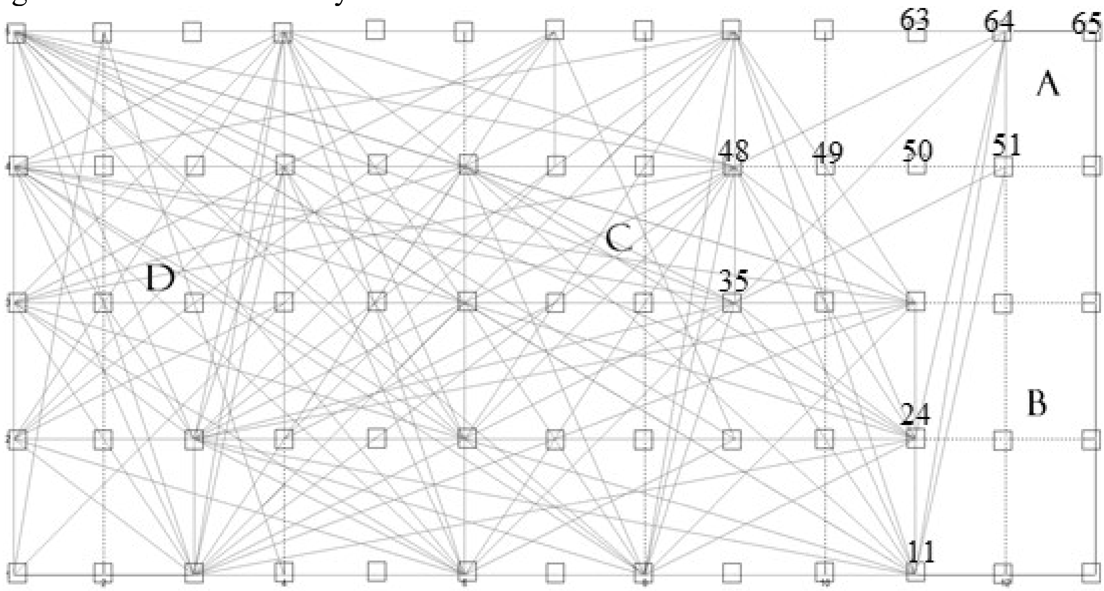


Figure 10 is a 2 column figure

Figure 10: Euclidean distances with values below 0.2.

Every dot represents the cells (hexagons) from map. If we start to number them from bottom-left, to right, we can find cell number 11 and others cells, written in the graph. We can observe the connectivity and inner relations between nodes based on Euclidean distances. Each straight line represents the connectivity between two cells. We can observe that Group D exhibits the highest level of connectivity.



Note: Lines indicate the distances between cells. Units based on Euclidean distances with values below 0.2.