Expert Systems With Applications 42 (2015) 8393–8402

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

(Dis)integration levels across global stock markets: A multidimensional scaling and cluster analysis

Hamidreza Esmalifalak [a,*], Ali Irannezhad Ajirlou [b], Sahar Pordeli Behrouz [c], Maryam Esmalifalak [d]

[a] Department of Management and Economics, Science and Research Branch, IAU, Tehran, Iran
[b] Department of Management, Science and Research Branch, IAU, Arak, Iran
[c] Department of Electrical and Computer Engineering, East Azarbaijan Science and Research Branch, IAU, Tabriz, Iran
[d] Department of Computer science, Houston Community College, HCC, Houston, USA

ARTICLE INFO

Keywords:
(Dis)integration levels
Global stock market (GSM)
Multidimensional scaling (MDS)
Agglomerative Hierarchical Clustering (AHC)

ABSTRACT

Following the globalization of the economy, interests in examining the financial linkages of economies have been increased, especially in the wake of the global financial crisis. In light of the benefits of international portfolio diversification, it is not surprising that there has been a strong interest among researchers to capture presence and extent of global stock market GSM integration. This study addresses to the feasibility and benefits of two visual data interpretation methods, based on multidimensional scaling and cluster analysis, in GSM integration context. The explanatory power of applied methods has become stronger by putting both integration and disintegration (anomaly) detection strategies. Empirically, in the first phase, we detect (dis)integration levels and clusters among 41 indexes from 40 countries. In the second phase, we choose most integrated indexes, detected in the first phase, to see whether their linkages are progressing, stable or regressing during different time series (12 seasons). In phase one and two, the topological properties of global stock networks are expressed in two-dimensional plots (dendrogram, heat map, frequency map). Based on the results, we found proposed methods capable of capturing (1) hierarchy of interrelated clusters embedded in a complex network of objects (2) different (dis)integration levels among selected global stock indexes and (3) abnormal seasons and indexes considering time and regional horizons.

1. Introduction

With the increasing significance of international trades and investments, integration of economic variables in different countries is becoming increasingly evident. In light of the benefits of international portfolio management, there is a tremendous interest among researchers to explain how global stock markets GSM are correlated with one and other. Most of forecasters in this field constantly search for predictable patterns to exploit trading opportunities. However, these attempts are not compatible with the efficient market hypothesis EMH. According to EMH, it is impossible to outperform the overall market through expert stock selection or market timing, and that the only way an investor can possibly obtain higher returns is by purchasing riskier investments. Nevertheless, in GSM transactions, we have many examples of integrated markets that their harmonization process is not fast enough, creating conditions for investors to have over-normal profits by making use of predefined patterns.

Due to the different degrees of international financial capital market efficiency, academic researchers investigate the EMH by exploring unknown and valuable knowledge from historical data (Liao & Chou, 2013). So far, in the analysis of stock market correlations, the ARCH-family models, especially Multivariate GARCH model, had prosperity among all the model-based analysis. However, more recently many researchers preferred to use model-free statistical testing and correlation analysis (Aityan, Ivanov-Schitz, & Izotov, 2010; Ilina & Daragan, 2001). From a methodological point of view, in all cases trading rules are generated and allow the investors to beat the market, confronting the EMH (Cervelló-Royo et al., 2015).

For some of recent researchers the dynamics of financial markets have attracted much attention. A new strand of literature has adopted the complex network theory to study the structure and topology of financial market integration. According to this theory, a financial market can be viewed as a complex system that consists of many individuals interacting with each other. Some have studied complex networks under random matrix theory RMT framework (Burda et al., 2011; Livan, Alfarano, & Scalas, 2011; Nobi,

* Corresponding author. Tel.: +18324550734.
E-mail addresses: huh_12@yahoo.com, esmalifalak.mohammad@gmail.com (H. Esmalifalak), aliajirlou@yahoo.com (A.I. Ajirlou), s.pordeli1890@yahoo.com (S.P. Behrouz), maryam.esmalifalak@student.hccs.edu (M. Esmalifalak).

http://dx.doi.org/10.1016/j.eswa.2015.06.053
0957-4174/

Maeng, Ha, & Lee, 2013; Wang, Xie, Chen, Yang, & Yang, 2013; Witte, 2014). Some others used the minimum spanning tree MST technique. Relevant works based on the MST include: in GSM indexes (Huang, Gu, Li, & Su, 2010), European Union stock market indexes (Gilmore, Lucey, & Boscia, 2008), Asia-pacific stock markets (Sensoy & Tabak, 2014) in global foreign exchange market (Naylor, Rose, & Moyle, 2007), individual stocks in the UK's stock market (Ulusoy et al., 2012) Germany's stock market (Brida & Risso, 2010), Turkey's stock market (Kantar, Deviren, & Keskin, 2011), China's stock market (Huang, Zhao, & Hou, 2009) and Brazil's stock market (Tabak, Serra, & Cajueiro, 2010).

In this paper we suggest methods called multidimensional scaling MDS and agglomerative hierarchical cluster analysis AHC to expand understanding GSM (dis)integration levels through visual inspection. Applied methods are the most common techniques for statistical data analysis and the main task of exploratory data mining. They can contribute to the definition of a formal classification scheme or taxonomy by revealing associations and structure in historical data which was not previously evident. Unlike the more traditional integration identification methods, these methods avoid imposing any specific restrictions over the tested items. In fact, the only required data is a measure for determining the (dis)integration between each possible pairs of objects. One of the basic limitations of these methods is that they are not clearly established consequently most of the guidelines for using these methods are rules of thumb. For instance, there are no completely satisfactory methods for determining the appropriate number of clusters.

1.1. Multidimensional scaling

Generally speaking MDS techniques develop spatial representations of psychological stimuli or other complex objects about which people make judgments (e.g., preference, (dis)similarity) (Carroll & Wish, 1975). In this method, the objects are represented as points in a map so that a small distance between two points corresponds to a high correlation (in this study integration) between two objects and a large distance corresponds to low or even negative correlation (in this study disintegration). Distance measure between the objects is generally easy to interpret and thus may be used to formulate more specific models or hypotheses. This method is most appropriate for achieving two objectives:

1. as an exploratory technique to identify unrecognized dimensions affecting behavior;
2. as a means of obtaining comparative evaluations of objects or sets of objects.

Researchers define a MDS analysis through three key decisions: selecting the objects that will be analyzed, deciding whether similarities or preference is to be analyzed and finally choosing whether the analysis will be performed at the group or individual level. In our study, we aim to analyze GSM indexes, we analyze markets integration or similarity (time-varying synchronization in market returns) and finally, we decide to perform analysis at both group and individual levels.

1.2. Cluster analysis

Most clustering routines, including the method used in this paper, is ubiquitously used in data mining as a method of discovering novel and actionable subsets within a set of data. Given a set of data *X*, the typical aim of partitional clustering is to form a k-block set partition Π_k of the data. In brief, clustering or (cluster analysis) is the task of grouping a set of objects in such an order that objects in the same group (cluster) are more similar to each other than to those in other groups (clusters). This notion of similarity can be expressed in very different ways, according to the purpose of the study, to domain-specific assumptions and to prior knowledge of the problem. In this study objects within a cluster regarded as integrated objects and others as outliers. In order to better understand the inter-cluster relations which embedded in a complex network of objects, we use Agglomerative Hierarchical Clustering AHC, a widely used method of building a cluster hierarchy or, in other words, a tree of clusters known as a Dendrogram[1] (see Figs. 2 and 5).

The outline of the paper is as follows; Section 2 briefly reviews various strands of the literature on stock market integrations. Section 3 presents the methodology of research. This methodology is then tested and analyzed for performance on the GSMs in Section 4. Our test is based on two phases. In the first phase, we conduct our empirical study between 41 countries indexes (region based analysis). Then, in the second phase we use the most integrated indexes detected in the first step to find any linkage in hierarchy of time frame clusters (Period based analysis). Finally, Section 5 presents some concluding remarks.

2. (Dis)integration across global stock markets: related studies

The overview of studies over the past decade indicates that the ARCH model and its modified types are the most popular models among the econometrics models to investigate integrations among stock market indexes. The ARCH model proposed by (Engle, 1982) and its modification Generalized-ARCH (GARCH) proposed by Bollerslev (1986). For a survey of ARCH-type models, see Bera and Higgins (1993), Bollerslev, Chou, and Kroner (1992), Bollerslev, Engle, and Nelson (1994), Pagan (1996), Palm (1996) and Bauwens, Laurent, and Rombouts (2006) among others. Although most of traditional studies have adopted the ARCH family models, relatively more new papers have also considered model-free data interpretation methodology. One of the approaches for detecting the integration is clustering approaches. General references regarding cluster analysis include works by Arabie, Hubert, and de Soete (1996), Dubes (1993), Everitt (1993), Fasulo (1999), Ghosh and Karin (2002), Hartigan (1975), Jain and Dubes (1988), Kaufman and Rousseeuw (1990), Spath (1980), Xu and Wunsch (2005) and works related to stock market including Aghabozorgi and Teh (2014), Arratia and Cabaña (2013), Chandrakala and Sekhar (2008), Durante and Foscolo (2013), Guo, Jia, and Zhang (2008), Liu, Cai, and Luo (2012), Nanda, Mahanty, and Tiwari (2010), Sandoval Junior (2013), Yu and Wang (2009). Here are some of related studies results:

Using tick-by tick transaction data, Ohta (2006) studies price clustering on the Tokyo stock market, (a computerized limit order market). According to their results, as for the intraday pattern, the degree of price clustering is maximal at the market opening and regressing during the first half hour until reaches a stable level. However, it does not increase again near the market closing. Basalto et al. (2005) applied a pairwise clustering approach to analyze Dow Jones index companies and identified similar temporal behavior in traded stock prices. They use a chaotic map clustering algorithm, where a map is identified to each company and the correlation coefficients of the financial time series to the coupling strengths between maps. In another study, Yu and Wang (2009) proposed a kernel base principal component analysis using *k*-means clustering algorithm to reduce the dimensionality of data and cluster stocks into different categories in terms of their financial information.

Boginski, Butenko, and Pardalos (2006) studied the market graph evolution over time. Interestingly, they conclude that the power-law structure of the market graph is quite stable over the considered time intervals. Liao and Chou (2013) employed association rules and clustering algorithm to study the linkage in the Taiwan and China (Hong Kong) stock markets. They categorized the stock indexes into thirty clusters to recognize the behavior of stock index associations. In their results, they provide different Patterns, rules, and clusters that are

[1] Dendrograms are a graphical representation of the taxonomic relationships of clusters produced by hierarchical clustering.