Physica A 450 (2016) 85–94

Contents lists available at ScienceDirect

Physica A

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

State and group dynamics of world stock market by principal component analysis

Ashadun Nobi[a,b], Jae Woo Lee[a,*]

[a] Department of Physics, Inha University, Incheon 402-751, South Korea

[b] Department of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Sonapur Noakhali-3802, Bangladesh

HIGHLIGHTS

- The principal component analysis (PCA) is applied to correlation coefficients of the world stock indices.
- The first PC shows a drastic change during the global financial crisis period.
- We observe the frequent change of PC coefficient in the American and Asian indices.
- The PCA is a valid tool to identify market states and to determine the subsets of global stock indices.

ARTICLE INFO

Article history:
Received 26 March 2015
Received in revised form 7 November 2015
Available online 14 January 2016

Keywords:
Principal component analysis
Stock market
Global financial crisis

ABSTRACT

We study the dynamic interactions and structural changes by a principal component analysis (PCA) to cross-correlation coefficients of global financial indices in the years 1998–2012. The variances explained by the first PC increase with time and show a drastic change during the crisis. A sharp change in PC coefficient implies a transition of market state, a situation which occurs frequently in the American and Asian indices. However, the European indices remain stable over time. Using the first two PC coefficients, we identify indices that are similar and more strongly correlated than the others. We observe that the European indices form a robust group over the observation period. The dynamics of the individual indices within the group increase in similarity with time, and the dynamics of indices are more similar during the crises. Furthermore, the group formation of indices changes position in two-dimensional spaces due to crises. Finally, after a financial crisis, the difference of PCs between the European and American indices narrows.

1. Introduction

Scientists from many fields have been attempting to understand the dynamics of financial markets for the past two decades [1–7]. There are many reasons for wanting to understand correlations in price behaviors. Due to estimated financial risk, statistical dependencies between stocks are of particular interest. Statistical dependencies within the market change with time due to the non-stationary behavior of the markets, which complicates the analysis. As a result, different kinds of methods and techniques are applied to analyze financial systems and extract its contained information [7–12]. Principal component analysis (PCA) is one of the established methods to characterize the evolving correlation structures of financial markets and to measure the associated systemic risk [13–16]. Generally, PCA is a multivariate statistical technique

* Corresponding author.
E-mail address: jaewlee@inha.ac.kr (J.W. Lee).

http://dx.doi.org/10.1016/j.physa.2015.12.144

particularly useful for analyzing the patterns of complex and multidimensional relationships that transform a large number of related observable variables into a smaller set of observable composite dimensions that can be used to represent their interrelationships. This method is used in different branches of science such as engineering, chemistry, and food technology to reduce the large dimensionality of the data sets and to characterize systems. In finance, most studies are carried out on financial sectors such as banks, brokers, insurance companies or hedge funds in order to measure systemic risks, arbitrage pricing theory and portfolio theory [17–24]. A recent study applied the PCA method to 10 different Dow Jones economic sector indices and showed that the larger was the change in PC1, the greater was the increase in systemic risk [16]. Here, we apply the PCA method to global financial indices to identify the market state of each index and to classify the groups of indices. The returns of some markets are particularly associated with groups of nations, and the returns of each one are based on the returns of the associated group. The application of the PCA method to global financial indices has been successfully performed [14,24]. In reference 14, the authors estimate the global factor as the first component using principal component analysis. In reference 24, the authors use a diverse range of asset classes such as equity indices, bonds, commodities, metals, currencies, etc. and consider weekly time series. They investigate the correlations using random matrix theory and the PCA method and identify the notable changes in assets during the credit and liquidity crisis in 2007–2008. In another recent study, the PCA method is applied to financial indicators of Europe, Japan and the United States, and the group of indicators is identified [25]. In our work, we use daily closing prices of 25 global indices from 1998 to 2012, analyzed in one-year time windows, and identify significant changes in market state due to external or internal crises over the entire period. Generally, different kinds of network techniques such as the minimum spanning tree, hierarchical method, planar maximal graph, and threshold method are applied to segment global equity markets [26,27].

For segmentation of global indices, we use the components of the first two PCs which is an innovative approach for complex non-stationary systems. We apply the load plots to observe the change of the state in the stock market when the crisis is approaching. The loading plots of the first three PC coefficients are calculated from correlations between principal components and returns (assets).

The rest of the paper is organized as follows: the financial data are discussed in Section 2. The method of PCA is explained in Section 3. The market state is analyzed in Section 4. The group dynamics are discussed in Section 5. Finally, we draw our conclusions in the final section.

2. Data analysis

We analyze the daily closing prices of 25 global indices from January 2, 1998 to December 20, 2012. These global financial indices are as follows: 1. Argentina (ARG), 2. Austria (AUT), 3. Australia (AUS), 4. Brazil (BRA), 5. Germany (GER), 6. India (IND), 7. Indonesia (INDO), 8. Israel (ISR), 9. South Korea (SKOR), 10. Malaysia (MAL), 11. Mexico (MEX), 12. The Netherlands (NETH), 13. Norway (NOR), 14. United Kingdom (UK), 15. United States (US), 16. Belgium (BEL), 17. Canada (CAN), 18. China (CHN), 19. France (FRA), 20. Hong Kong (HKG), 21. Japan (JPN), 22. Singapore, 23. Spain (ESPN), 24. Switzerland (SWZ), and 25. Taiwan (TWN). The sequence numbers (1, . . . , 25) and the abbreviations (ARG, . . . , TWN) of the indices above are used to label the indices in the pictures. The data are collected from Ref. [28]. To make an equal-time cross-correlation matrix; we remove some days on the basis of public holidays. If 30% of the markets are not open on a specific day, we remove that day and when it was fewer than 30%, we kept existing indices and inserted the last closing price for each of the reaming indices [26]. Thus, we considered all indices at the same date and filtered the data as in Ref. [29]. We examine daily returns for the indices, each containing approximately 260 records for each year. When we consider the correlation among the stock time series, the markets around the world do not operate at the same time. The time series of the stock market are not truly synchronous, even though we use the daily data [30–33]. Correlations of the S&P 500 with the other indices were stronger for indices that have overlapping working hours with the NYSE, and most indices that have no overlapping hours with the NYSE usually have larger correlations with the S&P 500 of the previous day [33]. In this work we consider the correlation of the stock time series on the same day. This should underestimate the correlation between the indices.

3. Principal component analysis in the stock market

We analyze the daily logarithmic return, which is defined for index i as

$$r_i(t) = \ln[I_i(t)] - \ln[I_i(t-1)], \tag{1}$$

where $I_i(t)$ is the closing price of index i on day t. The normalized return for index i is defined as

$$r'_i(t) = (r_i(t) - \langle r_i \rangle)/\sigma_i, \tag{2}$$

where $\sigma_i(r_i) = \sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2}$ is the standard deviation of the stock index time series i over the time window, and the symbol $\langle \cdots \rangle$ denotes average over the time window. Then, the normalized returns matrix is constructed from the time series r'_i with the dimensions $N \times T$ where $N = 25$ and $T = 260$.

$$R = \begin{bmatrix} r'_{1t1} & \cdots & r'_{1T} \\ \vdots & \ddots & \vdots \\ r'_{Nt1} & \cdots & r'_{NT} \end{bmatrix}. \tag{3}$$