Physica A 392 (2013) 3715–3730

Contents lists available at SciVerse ScienceDirect

Physica A

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

Random matrix theory analysis of cross-correlations in the US stock market: Evidence from Pearson's correlation coefficient and detrended cross-correlation coefficient

Gang-Jin Wang [a], Chi Xie [a,b,*], Shou Chen [a,b], Jiao-Jiao Yang [a], Ming-Yan Yang [a]

[a] *College of Business Administration, Hunan University, Changsha 410082, China*
[b] *Center of Finance and Investment Management, Hunan University, Changsha 410082, China*

HIGHLIGHTS

- Statistical properties of cross-correlations in the US stock market are analyzed by RMT.
- Statistical results of cross-correlations are showed from PCC and DCCA coefficient.
- New results of the cross-correlations in the US stock market are found.
- Interesting properties at different time scales are found by DCCA coefficient.

ARTICLE INFO

Article history:
Received 15 October 2012
Received in revised form 14 March 2013
Available online 24 April 2013

Keywords:
Econophysics
Random matrix theory
Cross-correlations
Detrended Cross-correlation coefficient
Pearson's correlation coefficient
US stock market

ABSTRACT

In this study, we first build two empirical cross-correlation matrices in the US stock market by two different methods, namely the Pearson's correlation coefficient and the detrended cross-correlation coefficient (DCCA coefficient). Then, combining the two matrices with the method of random matrix theory (RMT), we mainly investigate the statistical properties of cross-correlations in the US stock market. We choose the daily closing prices of 462 constituent stocks of S&P 500 index as the research objects and select the sample data from January 3, 2005 to August 31, 2012. In the empirical analysis, we examine the statistical properties of cross-correlation coefficients, the distribution of eigenvalues, the distribution of eigenvector components, and the inverse participation ratio. From the two methods, we find some new results of the cross-correlations in the US stock market in our study, which are different from the conclusions reached by previous studies. The empirical cross-correlation matrices constructed by the DCCA coefficient show several interesting properties at different time scales in the US stock market, which are useful to the risk management and optimal portfolio selection, especially to the diversity of the asset portfolio. It will be an interesting and meaningful work to find the theoretical eigenvalue distribution of a completely random matrix $\mathbf{R}$ for the DCCA coefficient because it does not obey the Marčenko–Pastur distribution.

1. Introduction

Random matrix theory (RMT) is a popular technical tool for investigating the cross-correlations in financial markets [1–3]. RMT, which was first proposed by Wigner [4] in 1951, can be used to describe the statistical properties of energy levels in

---

[*] Corresponding author at: College of Business Administration, Hunan University, Changsha 410082, China. Tel.: +86 731 88823890; fax: +86 731 88823670.

E-mail address: xiechi@hnu.edu.cn (C. Xie).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.04.027

complex quantum systems. In 1999, Laloux et al. [5] and Plerou et al. [6] applied the RMT approach to analyze the cross-correlations in the US stock market. The former chose the daily prices of 406 assets of the Standard & Poor (S&P) 500 index from 1991 to 1996 as the empirical data [5], and the research object of the latter was the 30-min high-frequency data of the largest 1000 US companies during the years 1994–1995 [6]. Although the empirical data of Refs. [5,6] are different, Laloux et al. [5] and Plerou et al. [6] came to a similar conclusion: the largest eigenvalue corresponding to the financial market (i.e., the US stock market), is roughly 25 times greater than the maximum eigenvalue (i.e., λ_+) predicted by RMT. The same results were drawn by Plerou et al. [7] in 2002 who studied three different empirical datasets: the first one is the same as Ref. [6], the second one is 30-min high-frequency data of 881 US stocks in the period of 1996–1997, and the third one is 1-day data of 422 US stocks over the 35 years period 1962–1996. However, Utsugi et al. [8] analyzed the daily prices of 297 stocks of S&P 500 index from January 1991 to July 2001, and found that the largest eigenvalue (=52.2) is approximately 29 times larger than the maximum eigenvalue ($\lambda_+ = 1.79$). Here, we may ask what happened in the US stock market in recent years which burst the US sub-prime crisis and the European debt crisis.

Since RMT was introduced in financial markets by Laloux et al. [5] and Plerou et al. [6], it has been widely used to study the statistical properties of cross-correlations in different financial markets [9–28], such as the Warsaw stock market [13,22], the South African stock market [15], the Istanbul stock market [16], the foreign exchange market [19], and the Indian stock market [20,21]. More recently, Eom et al. [23] examined the topological properties of stock networks by using the minimal spanning tree (MST) method and RMT in financial data. They stated that the largest eigenvalue evidently has an effect on the construction of stock networks. Namaki et al. [25] studied the stability of correlation matrices of both the Tehran stock market (as an emerging market) and the Dow Jones Industrial Average (DJIA, as a mature market) via the mean values of the distribution of correlation coefficients and RMT under the global perturbation. Their results showed that DJIA is more sensitive to perturbations than the Tehran stock market. Oh et al. [26] analyzed the statistical properties of the cross-correlations in the Korean stock market through RMT and found that the largest eigenvalue is 52 times greater than the maximum eigenvalue predicted by RMT. Kumar et al. [27] applied RMT to investigate the cross-correlation properties of 20 financial indices before and during the US sub-prime crisis. They observed that few of the largest eigenvalues deviate dramatically from the prediction by RMT and the deviation has changed during the US sub-prime crisis. The time-lag RMT (TLRMT) was also introduced in financial markets to investigate long-range collective movements [29,30]. For instance, Podobnik et al. [29] examined 88 constituent stocks of S&P 500 index in 2009 during the period 1983–2009 and revealed pronounced peaks in times of crisis by applying the TLRMT. Accurately, they showed singular values versus year and found marked peaks during the largest market shocks and financial crises, such as the Black Monday, the Dot-com Bubble, and the US sub-prime crisis. Wang et al. [30] investigated 48 world financial indices and employed the TLRMT to demonstrate the decay of cross-correlations with time lags. They reported long-range power-law cross-correlations in the volatilities (i.e., the absolute returns) which can quantify risk, and indicated that cross-correlations between the volatilities decay much more slowly than cross-correlations between the returns.

In the aforementioned previous works, the cross-correlation matrices are constructed by Pearson's correlation coefficient (PCC) C_{ij}, which is defined by between stock i and stock j, $C_{ij} = \frac{\langle r_i r_j \rangle - \langle r_i \rangle \langle r_j \rangle}{\sigma_i \sigma_j}$, where r_i and r_j are returns of stocks i and j respectively, $\sigma_i = \sqrt{\langle r_i^2 \rangle - \langle r_i \rangle^2}$ is the standard deviation of r_i and $\langle \cdots \rangle$ represents the time average over the period studied. PCC can describe the linear correlation between two time series which are both assumed to be stationary [31]. However, in the real world, especially in the financial markets, the time series are often non-stationary and heterogeneous [32,33]. In other words, PCC may lose effectiveness if the time series are non-stationary or non-Gaussian distributions. Zebende [34] also pointed out PCC is not robust and can be misleading if outliers are present because the real world recordings are characterized by a high level of heterogeneity. To address the drawbacks of PCC, based on the detrended fluctuation analysis (DFA) method [35] and the detrended cross-correlation analysis (DCCA) method [36,37], Zebende [34] recently developed a novel detrended cross-correlation coefficient, i.e., DCCA coefficient $\rho_{ij}(s)$, to quantify the level of cross-correlation between non-stationary time series, where s is the time scale. The outstanding advantage of this nonlinear cross-correlation coefficient is that it can investigate the cross-correlations at different time scales [38,39]. Vassoler and Zebende [40] adopted the DCCA coefficient to analyze and quantify cross-correlations between time series of air temperature and relative humidity. Podobnik et al. [41] studied the statistical significance of $\rho_{ij}(s)$, and showed that the tendency of the Chinese stock market to follow the US stock market is extremely weak by using the coefficient $\rho_{ij}(s)$. Taking into account what has been discussed above, we hereby put forward three questions as follows:

(i) Can the DCCA coefficient $\rho_{ij}(s)$ efficiently measure the cross-correlations of financial time series by combining RMT?
(ii) If the cross-correlations are calculated by PCC and $\rho_{ij}(s)$ respectively, what are the differences between the two coefficients?
(iii) What properties do the cross-correlations in financial markets have at different time scales, where the cross-correlations are measured by $\rho_{ij}(s)$ and RMT?

Therefore, in this study, we aim to answer the above-mentioned questions. In the empirical process, we first choose the daily closing prices of 462 constituent stocks of S&P 500 index from January 3, 2005 to August 31, 2012 as the research data. Next, we respectively use PCC and the DCCA coefficient $\rho_{ij}(s)$ to construct the empirical cross-correlation matrices of the 462 stocks and combine them with RMT to analyze the statistical properties of the cross-correlations in the US stock market.