Cross-correlation in financial dynamics

J. Shen and B. Zheng

Zhejiang University, Zhejiang Institute of Modern Physics - Hangzhou 310027, PRC

received 20 December 2008; accepted in final form 5 May 2009
published online 4 June 2009

PACS 89.65.Gh – Economics; econophysics, financial markets, business and management
PACS 89.75.-k – Complex systems

Abstract – To investigate the universal structure of interactions in financial dynamics, we analyze the cross-correlation matrix $C$ of price returns of the Chinese stock market, in comparison with those of the American and Indian stock markets. As an important emerging market, the Chinese market exhibits much stronger correlations than the developed markets. In the Chinese market, the interactions between the stocks in a same business sector are weak, while extra interactions in unusual sectors are detected. Using a variation of the two-factor model, we simulate the interactions in financial markets.

In recent years, there has been a growing interest of physicists in economic systems. Concepts and methods in physics have been applied to the study of financial time series [1–7]. Different models and theoretical approaches have been developed to describe the features of the financial dynamics [8–20]. Statistical properties of price fluctuations and correlations between different stocks are topics of interest, not only scientifically for understanding the complex structure and dynamics of the economy, but also practically for the asset allocation and portfolio risk estimation [21–23]. The probability distributions of stock prices in different stock markets show a universal nature and follow the “inverse cubic law” [24–26]. However, the statistical properties of correlations between different stocks seem less universal across different stock markets [27].

Unlike most traditional physical systems, where one derives correlations between subunits from their interactions, the underlying “interactions” for the stock markets are not known. Pioneering studies at the phenomenological level analyze cross-correlations between stocks by applying concepts and methods of the random matrix theory (RMT), which was developed in the context of complex quantum systems where the precise nature of the interactions between subunits is not known [28,29]. The properties of the empirical correlation matrix $C$ of price returns are compared with those of a random matrix in which the price movements are uncorrelated [30,31]. This spectral property-focused method was first applied to developed markets such as the New York Stock Exchange (NYSE) in USA [30–33], and recently also to some emerging markets, e.g. the National Stock Exchange (NSE) in India [27].

In general, the bulk of the eigenvalue spectrum of the correlation matrix $C$ of price returns shares universal properties with the Gaussian orthogonal ensemble of random matrices, while the largest eigenvalue of $C$ which deviates significantly from the bulk represents the influence of the entire market on all stocks. This is true for both developed and emerging stock markets. For developed markets, other large eigenvalues can be associated with the conventionally identified business sectors, according to the compositions of the eigenvectors. These sectors are stable in time, in some cases, as long as 30 years [32,34]. But for emerging markets such as the NSE in India, the number of large eigenvalues are smaller than that of developed markets [27]. Although there exist often much more correlations between stocks in emerging markets in all, the less large eigenvalues may reflect less correlations between stocks in a same sector. It seems that strong interactions may emerge within groups of stocks as a financial market evolves over time.

Although there have been many studies of correlated price movements in stock markets [35–39], the researches on emerging markets are few and mostly on the synchronicity of price movements across stocks [40–43]. The Chinese stock market is an important emerging market. Although sharing certain common features with developed markets, it also exhibits unusual dynamic properties [7,44–46]. Especially, the positive return-volatility correlation in the Chinese market is unique according to

E-mail: zheng@zimp.zju.edu.cn
up-to-date literatures [7,46]. In some sense, the Chinese market is a representative emerging market, and the Indian market is gradually approaching a developed one.

In this paper, we continue our comparative study of the Chinese market and developed markets as well as even the Indian market, with the RMT theory and using the data of the Shanghai Stock Exchange (SSE) of China. The Shanghai Stock Exchange was established in 1990, and it is now one of the most important emerging stock markets in the world. We have analyzed the daily closing prices of 259 stocks traded in the SSE from Jan., 1997 to Nov., 2007, and it corresponds to 2633 days. These data are obtained from “Wind Financial Database” (http://www.wind.com.cn), and those stocks which existed only for short periods are not included in our analysis. If the price of a stock is missing on a particular day, it is then assumed that the price remains the same as the preceding day [43]. In ref. [27], it has been testified that the missing prices do not result in artifacts for the American market.

Let \( P_i(t) \) be the price of a stock \( i = 1, \ldots, N \) at time \( t \), we then define the logarithmic price return of the \( i \)-th stock over a time interval \( \Delta t \) as

\[
R_i(t, \Delta t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t).
\]

We first set \( \Delta t \) to be one day. Since different stocks may fluctuate at different levels, e.g., measured by the standard deviation of its return, we should make sure that the results are independent of the scale of the measurement. Therefore, we introduce the normalized return

\[
r_i(t, \Delta t) \equiv \frac{R_i(t, \Delta t)}{\sigma_i},
\]

where \( \sigma_i \equiv \sqrt{(R_i^2) - \langle R_i \rangle^2} \) is the standard deviation of \( R_i \), and \( \langle \ldots \rangle \) represents the time average. Finally, we compute the equal-time cross-correlation matrix \( C \), whose element

\[
C_{ij} \equiv \langle r_i r_j \rangle
\]

measures the correlation between the returns of stocks \( i \) and \( j \). By definition, \( C \) is a real symmetric matrix with \( C_{ii} = 1 \), and \( C_{ij} \) is valued in the domain \([-1, 1]\).

In fig. 1(a), the probability distribution \( P(C_{ij}) \) of the elements of the cross-correlation matrix \( C \) is displayed for the SSE (China), NSE (India) and NYSE (USA). The data for the NSE and NYSE are from the period 1996–2006 [27]. The mean value \( \langle C_{ij} \rangle \) of the elements for the SSE is 0.37, much larger than 0.22 and 0.20 for the NSE and NYSE respectively [27]. This result strongly supports the general belief that stock prices in emerging markets tend to be more correlated than developed ones [40,47]. In this sense, however, the Chinese market is much more “emerging” than the Indian market. In addition, all elements of the matrix \( C \) of the SSE are positive, while there are a few negative ones for the NSE and NYSE. This also suggests that the correlation between stocks is stronger in the Chinese stock market.

If \( N \) time series of length \( T \) are mutually uncorrelated, the resulting cross-correlation matrix is called a Wishart matrix. Statistical properties of such random matrices are known [48,49]. In the limit \( N \to \infty \) and \( T \to \infty \) with \( Q \equiv T/N \geq 1 \), the probability distribution \( P_{\text{ran}}(\lambda) \) of the eigenvalue \( \lambda \) is given by

\[
P_{\text{ran}}(\lambda) = \frac{Q}{2\pi} \sqrt{(\lambda_{\text{ran}}^\max - \lambda)(\lambda - \lambda_{\text{ran}}^\min)}
\]

for \( \lambda_{\text{ran}}^\min \leq \lambda \leq \lambda_{\text{ran}}^\max \) and 0 otherwise. The lower (upper) bound is given by

\[
\lambda_{\text{ran}}^\min(\max) = [1 \pm (1/\sqrt{Q})]^2.
\]

For the data of the SSE in the period 1997–2007, for example, there are \( N=259 \) stocks each containing \( T=2633 \) returns. If there are no correlations between
the stocks, the eigenvalues should be bounded between \( \lambda_{\text{min}} \) and \( \lambda_{\text{max}} \). In developed stock markets such as those of USA and Japan, the bulk of the eigenvalue spectrum \( P(\lambda) \) of the cross-correlation matrix is similar to \( P_{\text{ran}}(\lambda) \) of the Wishart matrix, but some large eigenvalues deviate significantly from the upper bound \( \lambda_{\text{max}} \). This is also true for emerging stock markets, but the number of the large eigenvalues is relatively few as it is reported for the NSE (India) [27]. The latter should be a characteristic of emerging markets. In fig. 1(b), the eigenvalue spectrum of the SSE (China) is displayed with the solid line. The bulk is similar to \( P_{\text{ran}}(\lambda) \) of the Wishart matrix, and there are a few large eigenvalues. The number of the large eigenvalues is less than that of the NYSE (USA). To verify that these outliers are not an artifact of the finite length of the observation period, we shuffle the empirical time series of returns (i.e., randomly change the time sequence of returns), thereby destroying all the equal-time correlations, then compute a surrogate correlation matrix. The presence of a well-defined bulk of the eigenvalue spectrum which agrees with \( P_{\text{ran}}(\lambda) \) suggests that the contents of \( C \) are mostly random except for the large eigenvalues that deviate. As shown in table 1 and fig. 1(b), the largest eigenvalue \( \lambda_0 \) of the cross-correlation matrix of the SSE (China) is 97.33, about 56 times as large as the upper bound \( \lambda_{\text{max}} \) of \( P_{\text{ran}}(\lambda) \). However, \( \lambda_0 \) of the NSE (Indian) and NYSE (USA) is 46.67 and 42.50, only about 26 and 28 times as large as \( \lambda_{\text{max}} \), respectively. The correlation corresponding to the largest eigenvalue \( \lambda_0 \) is believed to be generated by interactions common for stocks in the entire market [27,30–33]. Our results imply that this correlation of the market mode in the Chinese stock market is much stronger than that in other stock markets analyzed.

Now let us look at the components \( u_i(\lambda) \) of the eigenvectors of the first four largest eigenvalues. In the top sector of fig. 2(a), the eigenvector of the largest eigenvalue \( \lambda_0 \) shows a relatively uniform composition with all stocks participating in it. Actually, all the elements also have the same sign. It represents a common mechanism that affects all the stocks with the same bias. So the largest eigenvalue is associated with the market mode, i.e., the collective response of the entire market to the external information. The Chinese market shares this feature with other stock markets [27,30,32,33]. According to previous studies [32,34], the components of the eigenvectors of other large eigenvalues are localized, i.e., each of these eigenvectors is dominated by only a part of stocks. The dominating components of an eigenvector belong to similar or related business sectors. This is true for both developed markets and some emerging markets [27,32,34]. However, the number of the large eigenvalues of the NYSE (USA) is obviously larger than that of the NSE (India) [27,32,34]. In fact, this number of

Table 1: The values of \( T, N, Q, \lambda_{\text{ran(min(max))}} \) and \( \lambda_{\text{real(min(max))}} \) of the SSE (China), in comparison with those of the NSE (India) and NYSE (USA) [27]. \( \lambda_{\text{ran(min(max))}} \) represents the low (upper) bound of the eigenvalues of the Wishart matrix, while \( \lambda_{\text{real(min(max))}} \) are those of real stock markets.

| Stock       | \( T \) | \( N \) | \( Q \) | \( \lambda_{\text{ran(min(max))}} \) | \( \lambda_{\text{real(min(max))}} \) |
|-------------|--------|--------|--------|-----------------------------------|-----------------------------------|
| China       | 2632   | 259    | 10.2   | 0.47                              | 1.73                              |
| India       | 2621   | 201    | 13.0   | 0.52                              | 1.63                              |
| USA         | 2606   | 201    | 13.0   | 0.52                              | 1.63                              |

Cross-correlation in financial dynamics
the SSE (China) is even smaller. Especially, the dynamic effect of business sectors is hardly seen, as shown in the second, third and fourth sectors of fig. 2(a).

**What interactions do control the large eigenvalues of the SSE?** To answer it, we introduce a threshold $u_c$ to select the dominating components of the eigenvectors with large eigenvalues. In other words, we consider that only those components which satisfy $|u_i(\lambda)| \geq u_c$ contribute to the eigenvector of the eigenvalue $\lambda$. The results of $u_c = 0.08$ is shown in fig. 2(b).

In the SSE (China), a company will be specially treated when its financial situation is abnormal. In this case, the acronym “ST” will then be added to the stock name as a prefix. The abnormal financial situation includes: the audited profits are negative in two successive accounting years, the audited net worth per share is less than its stock’s par value in the recent accounting year etc. When the company has a negative audited profit for three successive accounting years, the acronym “*ST” will be added to its stock name and the stock may be also delisted. When the abnormal financial conditions are eliminated, the acronym “ST” or “*ST” will be removed.

For the SSE (China), we observe that the eigenvector of the second largest eigenvalue $\lambda_1$ is dominated by the stocks now with or used to be with “ST” or “*ST”. When we increase the threshold $u_c$, the proportion of the dominating stocks with “ST” and “*ST” rises. When $u_c \geq 0.12$, all dominating components are “ST” or “*ST” stocks. Detailed results are given in table 2. Therefore, the second largest eigenvalue corresponds to the so-called ST sector.

The dominating components of the eigenvector of the third largest eigenvalue $\lambda_2$ mainly are or used to be “Blue-chip” stocks, which are referred to be with stable and good performance, i.e., a reasonable positive profit in a period of time. When we increase the threshold $u_c$, the proportion of the Blue-chip stocks steadily rises, as shown in table 2. With a similar approach, we examine the dominating components of the eigenvector of the forth largest eigenvalue $\lambda_3$, and observe that they are mainly those companies registered in Shanghai with the real estate business (SHRE). When we continue to investigate the next largest eigenvalues such as $\lambda_4$ and $\lambda_5$ of the SSE (China), we could not find clear characteristics. For the NYSE (USA), however, one could detect the dominating effect of the business sectors at least up to the 10th largest eigenvalue [32, 34].

**Why does the SSE (China) exhibit the usual sectors like the ST and Blue-chip sectors rather than the standard business sectors?** Since the Chinese stock market is an emerging market, the companies are not operated strictly with the registered business. On the other hand, people seriously look at the performance of the companies. The fourth largest eigenvalue of the SSE may reflect the fact that Shanghai and especially its real estate business play important roles in China in the past years both economically and politically.

To investigate the possible dependence of our conclusions on the time scale, we change $\Delta t$ to one week, i.e., 5 working days. Our careful analysis shows that the results qualitatively remain the same. Quantitatively, the correlations between stocks on the weekly scale are slightly stronger than those on the daily scale. Finally, one may wonder whether the cross-correlation matrix evolves with time. For this purpose, we divide the whole data of the SSE into two subsets, for example, those from 1997–2001 and 2002–2007. The results of each subset remain nearly the same as those of the whole data. If we consider the subset with a shorter period, e.g., from 2005–2007 or 2006–2007, it would be somewhat fluctuating. But we could still observe a weak trend, that the Chinese market tends to less emerging as time evolves. For example, the largest eigenvalue $\lambda_0$ is 93.67 for the period 2005–2007, compared to 97.40 for 1997–2001.

To better understand the structure of interactions in the SSE (China), we introduce a variation of the two-factor model of the market dynamics [27]. We assume that the normalized return of the $i$-th stock from the $k$-th business sector can be decomposed into i) a market factor $r_m(t)$, containing information common to all stocks; ii) a business-sector factor $r_k(t)$, reflecting dynamic effects in the $k$-th business sector; iii) a profit factor $r_p(t)$, reflecting three categories of stocks: the “ST” and “Blue-chip” sectors, and the rest stocks; iv) a random factor $\eta(t)$. We thus obtain

$$r_k(t) = \beta_m r_m(t) + \gamma_k r_k(t) + \gamma_p r_p(t) + \sigma \eta(t),$$

(6)

where the constants $\beta$, $\gamma_k$, $\gamma_p$ and $\sigma$ represent the relative strengths of the four factors. For simplicity, we choose $r_m(t)$, $r_k(t)$, $r_p(t)$ and $\eta(t)$ to be Gaussian processes with zero mean and unit variance. The unit variance of $r_k(t)$ ensures

$$\beta^2 + (\gamma_k)^2 + (\gamma_p)^2 + \sigma^2 = 1.$$  

(7)

For each stock, we can independently assign $\gamma_k$, $\gamma_p$, and $\beta$, and obtain $\beta$ from eq. (7). We practically choose $\gamma_k$, $\gamma_p$, and $\sigma$ from a uniform distribution with a width $\delta$ and centered about the mean value $\gamma$, $\sigma$, and $\gamma_p$, respectively.

With the above model, we generate $N$ times series of length $T$ for returns $r_k$. The $K$ business sectors are composed of $n_1$, $n_2$, ..., $n_K$ stocks. The “ST” and “Blue-chip” stocks are randomly selected, and their numbers are $n_{ST}$ and $n_{BC}$ respectively. To describe the dynamic properties of the SSE (China), we choose a large $\beta$, $\gamma$.
We choose 5 sectors A, B, C, D and E, each with 50 stocks.

If the business sectors correspond to the first four largest eigenvalues of the cross-correlation matrix C from numerical simulations. The stocks are separated by dashed lines according to business sectors. Additionally, we take all \( \gamma_j \) for the ST sector bigger than any \( \gamma_j \) for the Blue-chip sector. For example, we take \( N = 250 \) and \( T = 2500 \), and assign 5 sectors A, B, C, D and E, each with 50 stocks. We choose \( \gamma_1 = 0.2 \), \( \gamma_2 = 0.55 \) for the ST sector and \( \gamma_3 = 0.40 \) for the Blue-chip sector, and \( \sigma = 0.7 \), \( \delta = 0.05 \), \( n_{ST} = 40 \) and \( n_{BC} = 40 \). With these fixed parameters, one calculates \( \beta \) to be 0.41 and 0.56 for the ST and Blue-chip sectors, respectively. Thus we obtain the first eight largest eigenvalues 105.80, 11.60, 5.41, 2.67, 2.52, 2.43, 2.31 and 1.91, in comparison with 97.33, 4.17, 3.35, 2.87, 2.62, 2.26, 2.09 and 1.79 of the SSE (China).

In fig. 3, the components of the eigenvectors of the first eight largest eigenvectors are displayed. The largest eigenvalue represents the market mode, and the components are distributed with uniform weights. For the second and third eigenvalues, we observe large components from different business sectors. When we introduce a threshold \( u_c \) to select those dominating components which satisfy \( |u_i| \geq u_c \), we find that they are all ST stocks for the second largest eigenvalue and Blue-chip stocks for the third largest eigenvalue.

For the next largest eigenvalues, we observe that several business sectors may contribute to a single eigenvector, attributed to the extra interactions between the stocks in the unusual sectors such as the ST and Blue-chip sectors. This qualitatively explains the weak effect of the business sectors in the Chinese stock market. For example, the SHRE sector corresponding to \( \lambda_3 \) in fig. 2(b) is not a standard business sector, and it is rather a mixture of different business sectors. To draw fig. 3(a) exactly the same as fig. 2(b), we can randomly assign the stocks of sector B and C to other business sectors. But it is regretful that the sectors of \( \lambda \leq \lambda_4 \) could not be clearly identified.

In ref. [50], an alternative stochastic model has been proposed for the stock-stock correlations. Our model actually shares a similar spirit with it in understanding the structure of interactions between stocks. For example, our market factor \( \beta_i r_m(t) \) and business-sector factor \( \gamma^* k^* g^*(t) \) correspond to the interactions between stocks with the coupling constants \( \epsilon_M \) and \( \epsilon_g \), respectively, in ref. [50].

In summary, we phenomenologically compute the cross-correlation matrix \( C \) of stock price returns. The Chinese stock market generally exhibits much stronger correlations than the developed stock markets. On the other hand, the correlations between stocks in a standard business sector are rather weak in the Chinese market. However, we could identify the large eigenvalues of the cross-correlation matrix \( C \) of the Chinese market according to unusual sectors such as the ST and Blue-chip sectors. Our results show that the Chinese market is a representative emerging market in the world.

Fig. 3: (a) The absolute values of the components \( u_i(\lambda) \) of stock \( i \) corresponding to the first four largest eigenvalues of the cross-correlation matrix \( C \) from numerical simulations. The stocks are separated by dashed lines according to business sectors. (b) A similar figure as (a), but for next four largest eigenvalues of \( C \) from numerical simulations. \( \lambda_4 \) corresponds to the upper bound \( \lambda_{\text{max}}^{\max} \) of the Wishart matrix.

This work was supported in part by NNSF (China) under grant Nos. 10875102 and 10325520.

REFERENCES

[1] Mantegna R. N. and Stanley H. E., Nature, 376 (1995) 46.
[2] Gopikrishnan P., Plerou V., Amaral L. A. N., Meyer M. and Stanley H. E., Phys. Rev. E, 60 (1999) 5305.
[3] Liu Y., Gopikrishnan P., Cizeau P., Meyer M., Peng C. K. and Stanley H. E., Phys. Rev. E, 60 (1999) 1390.
[4] Bouchaud J. P., Matacz A. and Potters M., Phys. Rev. Lett., 87 (2001) 228701.
[5] Gabaix X., Gopikrishnan P., Plerou V. and Stanley H. E., Nature, 423 (2003) 267.
[6] Ren F. and Zheng B., Phys. Lett. A, 313 (2003) 312.
