PHYSICAL REVIEW E **83**, 046121 (2011)

Quantifying and modeling long-range cross correlations in multiple time series with applications to world stock indices

Duan Wang,[1] Boris Podobnik,[1,2,3] Davor Horvatić,[4] and H. Eugene Stanley[1]

[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[2]*Faculty of Civil Engineering, University of Rijeka, 51000 Rijeka, Croatia*
[3]*Faculty of Economics, University of Ljubljana, 1000 Ljubljana, Slovenia*
[4]*Physics Department, Faculty of Science, University of Zagreb, 10000 Zagreb, Croatia*

(Received 3 December 2010; published 25 April 2011)

We propose a modified time lag random matrix theory in order to study time-lag cross correlations in multiple time series. We apply the method to 48 world indices, one for each of 48 different countries. We find long-range power-law cross correlations in the absolute values of returns that quantify risk, and find that they decay much more slowly than cross correlations between the returns. The magnitude of the cross correlations constitutes "bad news" for international investment managers who may believe that risk is reduced by diversifying across countries. We find that when a market shock is transmitted around the world, the risk decays very slowly. We explain these time-lag cross correlations by introducing a global factor model (GFM) in which all index returns fluctuate in response to a single global factor. For each pair of individual time series of returns, the cross correlations between returns (or magnitudes) can be modeled with the autocorrelations of the global factor returns (or magnitudes). We estimate the global factor using principal component analysis, which minimizes the variance of the residuals after removing the global trend. Using random matrix theory, a significant fraction of the world index cross correlations can be explained by the global factor, which supports the utility of the GFM. We demonstrate applications of the GFM in forecasting risks at the world level, and in finding uncorrelated individual indices. We find ten indices that are practically uncorrelated with the global factor and with the remainder of the world indices, which is relevant information for world managers in reducing their portfolio risk. Finally, we argue that this general method can be applied to a wide range of phenomena in which time series are measured, ranging from seismology and physiology to atmospheric geophysics.

DOI: 10.1103/PhysRevE.83.046121 PACS number(s): 89.65.Gh, 89.20.−a, 02.50.Ey

I. INTRODUCTION

When complex systems join to form even more complex systems, the interaction of the constituent subsystems is highly random [1–4]. The complex stochastic interactions among these subsystems are commonly quantified by calculating the cross correlations. This method has been applied in systems ranging from nanodevices [5–7], atmospheric geophysics [8], and seismology [9–11], to finance [12–23]. Here we propose a method of estimating the most significant component in explaining long-range cross correlations.

Studying cross correlations in these diverse physical systems provides insight into the dynamics of natural systems and enables us to base our prediction of future outcomes on current information. In finance, we base our risk estimate on cross correlation matrices derived from asset and investment portfolios [15,24]. In seismology, cross correlation levels are used to predict earthquake probability and intensity [10]. In nanodevices used in quantum information processing, electronic entanglement necessitates the computation of noise cross correlations in order to determine whether the sign of the signal will be reversed when compared to standard devices [5]. In Ref. [25], cross correlations reported for $\Delta t = 0$ calculated between pairs of electroencephalogram (EEG) time series are inversely related to dissociative symptoms (psychometric measures) in 58 patients with paranoid schizophrenia. In genomics data, spatial cross correlations are reported [26] corresponding to a chromosomal distance of $\approx 10 \times 10^6$ base pairs. In physiology, a statistically significant difference is reported [26] between alcoholic and control subjects.

Many methods have been used to investigate cross correlations between pairs of simultaneously recorded time series [21,22] or among a large number of simultaneously recorded time series [15,27,28]. In Ref. [28], a power mapping of the elements is used in the correlation matrix that suppresses noise. In Ref. [21], the authors propose detrended cross correlation analysis, which is an extension of detrended fluctuation analysis [29] and is based on detrended covariance. In Ref. [22] a method is proposed for estimating the cross correlation function C_{xy} of long-range correlated series x_t and y_t. For fractional Brownian motions with Hurst exponents H_1 and H_2, the asymptotic expression for C_{xy} scales as a power of n with exponents H_1 and H_2.

Univariate (single) financial time series modeling has long been a popular technique in science. To model the autocorrelation of univariate time series, traditional time series models such as autoregressive moving average (ARMA) models have been proposed [30]. The ARMA model assumes that variances are constant with time. However, empirical studies accomplished on financial time series commonly show that variances change with time. To model time-varying variance, the autoregressive conditional heteroskedasticity (ARCH) model was proposed [31]. Since then, many extensions of ARCH have been proposed, including the generalized autoregressive conditional heteroskedasticity (GARCH) model [32] and the fractionally integrated autoregressive conditional heteroskedasticity model [33]. In these models, long-range autocorrelations in magnitudes exist, so a large price change at one observation is expected to be followed by a large price change at the next observation. Long-range autocorrelations in magnitude of signals have been reported in finance [33], physiology [34,35], river flow data [36], and weather data [37].

Besides univariate time series models, modeling correlations in multiple time series has been an important objective because of its practical importance in finance, especially in portfolio selection and risk management [38,39]. In order to capture potential cross correlations among different time series, models for coupled heteroskedastic time series have been introduced [40–42]. However, in practice, when those models are employed the number of parameters to be estimated can be quite large.

A number of researchers have applied multiple time series analysis to world indices, mainly in order to analyze zero-time-lag cross correlations. In Ref. [12], it was reported that for international stock returns of nine highly developed economies, the cross correlations between each pair of stock returns fluctuate strongly with time, and increase in periods of high market volatility. By volatility we mean time-dependent standard deviation of return. The finding that there is a link between zero-time-lag cross correlations and market volatility is "bad news" for global money managers who typically reduce their risk by diversifying stocks throughout the world. In order to determine whether financial crises are short lived or long lived, the authors of Ref. [43] recently reported that, for six Latin American markets, the effects of a financial crisis are short range. Between two and four months after each crisis, each Latin American market returns to a low-volatility regime.

In order to determine whether financial crisis are short term or long term at the world level, we study 48 world indices, one for each of 48 different countries. We analyze cross correlations among returns and magnitudes, for zero and nonzero time lags. We find that cross correlations between magnitudes last substantially longer than between the returns, similar to the properties of autocorrelations in stock market returns [44]. We propose a general method in order to extract the most important factors controlling cross correlations in time series. Based on random matrix theory (RMT) [15] and principal component analysis [27] we propose how to estimate the global factor and the most significant principal components in explaining the cross correlations. This method has the potential to be broadly applied in diverse phenomena where time series are measured, ranging from seismology to atmospheric geophysics.

This paper is organized as follows. In Sec. II we introduce the data analyzed and the definition of return and magnitude of return. In Sec. III we introduce a modified time-lag random matrix theory (TLRMT) to show the time-lag cross correlations between the returns and magnitudes of world indices. Empirical results show that the cross correlations between magnitudes decay more slowly than those between returns. In Sec. IV we introduce a single global factor model to explain the short- or long-range correlations among returns or magnitudes. The model relates the time-lag cross correlations among individual indices with the autocorrelation function of the global factor. In Sec. V we estimate the global factor by minimizing the variance of residuals using principal component analysis (PCA), and we show that the global factor does in fact account for a large percentage of the total variance using RMT. In Sec. VI we show the applications of the global factor model, including risk forecasting of the world economy, and finding countries that have the most independent economies.

II. DATA ANALYZED

In order to estimate the level of relationship between individual stock markets—both long-range and short-range cross correlations exist at the world level—we analyze $N = 48$ worldwide financial indices $S_{i,t}$ where $i = 1,2,\ldots,48$ denotes the financial index and t denotes the time. We analyze one index for each of 48 different countries: 25 European indices [45], 15 Asian indices (including Australia and New Zealand) [46], two American indices [47], and four African indices [48]. In studying 48 economies that include both developed and developing markets we significantly extend previous studies in which only developed economies were included (e.g., the seven economies analyzed in Refs. [13,49], and the 17 countries studied in Ref. [50]). We use daily stock index data taken from *Bloomberg*, as opposed to weekly [50] or monthly data [12]. The data cover the period 4 January 1999 through 10 July 2009, 2745 trading days. For each index $S_{i,t}$, we define the relative index change (return) as

$$R_{i,t} \equiv \ln S_{i,t} - \ln S_{i,t-1}, \tag{1}$$

where t denotes the time, in units of one day. By "magnitude of return" we denote the absolute value of return after removing the mean,

$$|r_{i,t}| \equiv |R_{i,t} - \langle R_{i,t}\rangle|. \tag{2}$$

III. MODIFIED TIME-LAG RANDOM MATRIX THEORY

A. Basic ideas of time-lag random matrix theory

In order to quantify the cross correlations, random matrix theory (Refs. [51,52]) was proposed in order to analyze collective phenomena in nuclear physics. In Ref. [15], the RMT was extended to cross correlation matrices in order to find cross correlations in collective behavior of financial time series. The largest eigenvalue λ_+ and smallest eigenvalue λ_- of the Wishart matrix W (a correlation matrix of uncorrelated time series with finite length) are

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}, \tag{3}$$

where $Q \equiv T/N$ (> 1), and N is the matrix dimension and T the length of each time series. The larger the discrepancy between (a) the correlation matrix C between empirical time series and (b) the Wishart matrix W obtained between uncorrelated time series, the stronger are the cross correlations in empirical data [15]. Many RMT studies reported equal-time (zero Δt) cross correlations between different empirical time series [15,53–56].

Recently time-lag generalizations of RMT have been proposed [57–59]. In one of the generalizations of RMT based on the eigenvalue spectrum called time-lag RMT, long-range cross correlations were found [26] in time series of price fluctuations in absolute values of 1340 members of the New York Stock Exchange Composite, in both healthy and pathological physiological time series, and in the mouse genome.

We compute for varying time lags Δt the largest singular values $\lambda_L(\Delta t)$ of the cross correlation matrix of the N-variable time series $X_{i,t}$,

$$C_{ij}(\Delta t) \equiv \frac{\langle X_{i,t} X_{j,t+\Delta t}\rangle - \langle X_{i,t}\rangle\langle X_{j,t+\Delta t}\rangle}{\sigma_i \sigma_j}. \tag{4}$$

We also compute $\tilde{\lambda}_L(\Delta t)$ of a similar matrix $\tilde{C}(\Delta t)$, where $X_{i,t}$ are replaced by the magnitudes $|X_{i,t} - \langle X_{i,t}\rangle|$. The squares of the nonzero singular values of C are equal to the nonzero eigenvalues of CC^+ or $\mathsf{C}^+\mathsf{C}$, where C^+ denotes the transpose of C. In a singular value decomposition (SVD) [26,59,60], $\mathsf{C} = \mathsf{UDV}^+$, the diagonal elements of D are equal to singular values of C, where the U and V correspond to the left and right singular vectors of the corresponding singular values. We apply SVD to the correlation matrix for each time lag and calculate the singular values. The dependence of the largest singular value $\lambda_L(\Delta t)$ on Δt serves to estimate the functional dependence of the collective behavior of C_{ij} on Δt [26].

B. Modifications of cross correlation matrices

We make two modifications of correlation matrices in order to better describe correlations for both zero and nonzero time lags.

(i) The first modification is a correction for correlation between indices that are not frequently traded. Since different countries have different holidays, all indices contain a large number of zeros in their returns. These zeros lead us to underestimate the magnitude of the correlations. To correct for this problem, we define a modified cross correlation between those time series with extraneous zeros,

$$C'_{ij}(\Delta t) \equiv \frac{1}{T'} \frac{\sum_{i=1}^{T} X_{i,t} X_{j,t+\Delta t} - \sum_{i=1}^{T} X_{i,t} \sum_{i=1}^{T} X_{j,t+\Delta t}}{\sigma_i \sigma_j}. \tag{5}$$

Here T' is the time period during which both $X_{i,t}$ and $X_{j,t+\Delta t}$ are nonzero. With this definition, the time periods during which $X_{i,t}$ or $X_{j,t+\Delta t}$ exhibit zero values have been removed from the calculation of cross correlations. The relationship between $C'_{ij}(\Delta t)$ and $C_{ij}(\Delta t)$ is

$$C'_{ij}(\Delta t) = \frac{T}{T'} C_{ij}(\Delta t). \tag{6}$$

(ii) The second modification corrects for autocorrelations. The main diagonal elements in the correlation matrix are 1's for zero-lag correlation matrices and autocorrelations for nonzero-lag correlation matrices. Thus, time-lag correlation matrices allow us to study both autocorrelations and time-lag cross correlations. If we study the decay of the largest singular value, we see a long-range decay pattern if there are long-range autocorrelations for some indices but no cross correlation between indices. To remove the influence of autocorrelations and isolate time-lag cross correlations, we replace the main diagonals by unity,

$$C''_{ij}(\Delta t) = \begin{cases} 1 & \text{when} \quad i = j, \\ C'_{ij}(\Delta t) & \text{when} \quad i \neq j. \end{cases} \tag{7}$$

With this definition the influence of autocorrelations is removed and the trace is kept the same as in the zero-time-lag correlation matrix.

C. Empirical results

In Fig. 1(a) we show the distribution of cross correlations between zero and nonzero lags. For $\Delta t = 0$ the empirical probability distribution function (PDF) $P(C_{ij})$ of the cross correlation coefficients C_{ij} substantially deviates from the corresponding PDF $P(W_{ij})$ of a Wishart matrix, implying the existence of equal-time cross correlations.

In order to determine whether short-range or long-range cross correlations accurately characterize world financial markets, we next analyze cross correlations for ($\Delta t \neq 0$). We find that with increasing Δt the form of $P(C_{ij})$ quickly

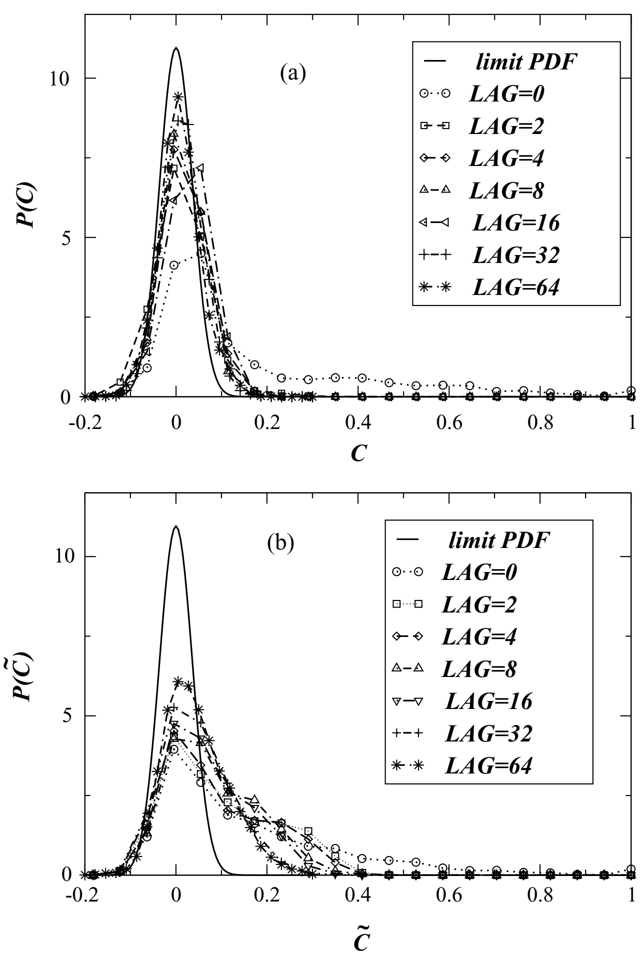

FIG. 1. Cross correlations among the $N = 48$ world financial index returns each of size $T = 2744$. (a) The empirical PDF of the coefficients of the cross correlation matrix C calculated between index returns with increasing Δt quickly converges to the Gaussian form. The normal distribution is the distribution of the pairwise cross correlations for finite length uncorrelated time series, which is a normal distribution with mean zero and standard deviation $\frac{1}{\sqrt{T}}$. (b) The empirical PDF of the coefficients of the matrix $\tilde{\mathsf{C}}$ calculated between index volatilities approaches the PDF of the random matrix more slowly than in (a).

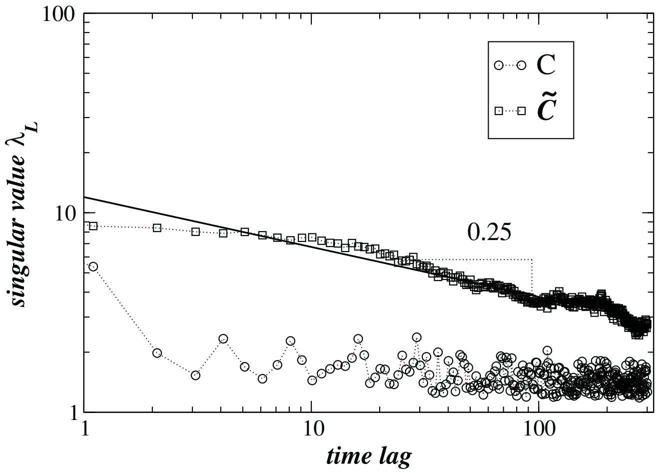

FIG. 2. Long-range magnitude cross correlations. The largest singular value λ_L obtained from the spectrum of the matrices C and $\tilde{C}$ versus time lag Δt. With increasing Δt, the largest singular values obtained for C of returns decays more quickly than $\tilde{C}$ calculated for absolute values of returns. The magnitude cross correlations decay as a power-law function with the scaling exponent of ≈0.25.

approaches the PDF $P(W_{ij})$, which is normally distributed with zero mean and standard deviation $1/\sqrt{N}$ [61].

In Fig. 1(b) we also show the distribution of cross correlations between *magnitudes*. In financial data, returns $R_{i,t}$ are generally uncorrelated or short-range autocorrelated, whereas the magnitudes are generally long-range autocorrelated [33,62]. We thus examine the cross correlations $\tilde{C}_{ij}(\Delta t)$ between $|r_{i,t}|$ for different Δt. In Fig. 1(b) we find that with increasing Δt, $P(\tilde{C}_{ij})$ approaches the PDF of random matrix $P(W_{ij})$ more slowly than $P(C_{ij})$, implying that cross correlations between index magnitudes persist longer than cross correlations between index returns.

In order to demonstrate the decay of cross correlations with time lags, we apply modified TLRMT. Figure 2 shows that with increasing Δt the largest singular value calculated for $\tilde{C}$ decays more slowly than the largest singular value calculated for **C**. This result implies that among world indices, the cross correlations between magnitudes last longer than cross correlations between returns. In Fig. 2 we find that λ_L vs Δt decays as a power-law function with the scaling exponent equal to 0.25. The faster decay of λ_L vs Δt for **C** implies very weak (or zero) cross correlations among world index returns for larger Δt, which agrees with the empirical finding that world indices are often uncorrelated in returns. Our findings of long-range cross correlations in magnitudes among the world indices is, besides a finding in Ref. [12], another piece of bad news for international investment managers. World market risk decays very slowly. Once the volatility (risk) is transmitted across the world, the risk lasts a long time.

IV. GLOBAL FACTOR MODEL

The arbitrage pricing theory states that asset returns follow a linear combination of various factors [63]. We find that the factor structure can also model time-lag pairwise cross correlations between the returns and between magnitudes. To simplify the structure, we model the time-lag cross correlations with the assumption that each individual index fluctuates in response to one common process, the global factor M_t,

$$R_{i,t} = \mu_i + b_i M_t + \epsilon_{i,t}. \tag{8}$$

Here, in the global factor model (GFM), μ_i is the average return for index i, M_t is the global factor, and $\epsilon_{i,t}$ is the linear regression residual, which is independent of M_t, with mean zero and standard deviation σ_i. Here b_i indicates the covariance between $R_{i,t}$ and M_t, $\mathrm{Cov}(R_{i,t},M_t) = b_i \mathrm{Var}(M_t)$. This single factor model is similar to the Sharpe market model [64], but instead of using a known financial index as the global factor M_t, we use factor analysis to find M_t, which we introduce in the next section. We also choose M_t as a zero-mean process, so the expected return $E(R_{i,t}) = \mu_i$, and the global factor M_t is only related with market risk. We define a zero-mean process $r_{i,t}$ as

$$r_{i,t} \equiv R_{i,t} - E(R_{i,t}) = b_i M_t + \epsilon_{i,t}. \tag{9}$$

A second assumption is that the global factor can account for most of the correlations. Therefore we can assume that there are no correlations between the residuals of each index, $\mathrm{Cov}(\epsilon_{i,t},\epsilon_{j,t}) = 0$. Then the covariance between $R_{i,t}$ and $R_{j,t}$ is

$$\mathrm{Cov}(R_{i,t},R_{j,t}) = \mathrm{Cov}(r_{i,t},r_{j,t}) = b_i b_j \mathrm{Var}(M_t). \tag{10}$$

The covariance between magnitudes of returns depends on the return distribution of M_t and $R_{i,t}$, but the covariance between squared magnitudes $r_{i,t}^2$ indicates the properties of the magnitude cross correlations. The covariance between $r_{i,t}^2$ and $r_{j,t}^2$ is

$$\mathrm{Cov}\left(r_{i,t}^2,r_{j,t}^2\right) = b_i^2 b_j^2 \mathrm{Var}\left(M_t^2\right). \tag{11}$$

The above results in Eqs. (10) and (11) show that the variance of the global factor and square of the global factor account for all the zero-time-lag covariance between returns and squared magnitudes. For time-lag covariance between $r_{i,t}$, we find

$$\mathrm{Cov}(r_{i,t},r_{j,t},\Delta t) = E(r_{i,t},r_{j,t-\Delta t}) - E(r_{i,t})E(r_{j,t-\Delta t}) \tag{12}$$
$$= b_i b_j A_M(\Delta t). \tag{13}$$

Here

$$A_M(\Delta t) \equiv E(M_t M_{t-\Delta t}) - E(M_t)E(M_{t-\Delta t}) \tag{14}$$

is the autocovariance of M_t. Similarly, we find

$$\mathrm{Cov}\left(r_{i,t}^2,r_{j,t}^2,\Delta t\right) = b_i^2 b_j^2 A_{M^2}(\Delta t). \tag{15}$$

Here

$$A_{M^2}(\Delta t) = E\left(M_t^2 M_{t-\Delta t}^2\right) - E\left(M_t^2\right)E\left(M_{t-\Delta t}^2\right) \tag{16}$$

is the autocovariance of M_t^2.

In the GFM, the time-lag covariance between each pair of indices is proportional to the autocovariance of the global factor. For example, if there is short-range autocovariance for M_t and long-range autocovariance for M_t^2, then for individual indices the cross covariance between returns will be short range and the cross covariance between magnitudes will be long range. Therefore, the properties of time-lag cross correlations in multiple time series can be modeled with a single time series—the global factor M_t.

The relationship between time-lag covariance among two index returns and autocovariance of the global factor also holds for the relationship between time-lag cross correlations among two index returns and autocorrelation function of the global factor, because it only needs to normalize the original time series to mean zero and standard deviation 1.

V. ESTIMATION AND ANALYSIS OF THE GLOBAL FACTOR

A. Estimation of the global factor

In contrast to domestic markets where, for a given country, we can choose the stock index as an estimator of the global factor, when we study world markets the global factor is unobservable. At the world level when we study cross correlations among world markets, we estimate the global factor using principal component analysis [27].

In this section we use bold font for N-dimensional vectors or $N \times N$ matrices, and subscript t for time series. Suppose $\mathbf{R}_t \equiv (R_{1,t}, R_{2,t}, \ldots, R_{N,t})^T$ is the multiple time series, each row of which is an individual time series $R_{i,t} = (R_{i,1}, R_{i,2}, \ldots, R_{i,T})$. We standardize each time series to zero mean and standard deviation 1 as

$$z_{i,t} \equiv \frac{R_{i,t} - \langle R_{i,t} \rangle}{\sigma(R_{i,t})}. \quad (17)$$

The correlation matrix can be calculated as $\mathbf{C} \equiv \frac{1}{T}\mathbf{z}_t\mathbf{z}_t^T$ where $\mathbf{z}_t^T$ is the transpose of $\mathbf{z}_t$, and the T in the denominator is the length of each time series. Then we diagonalize the $N \times N$ correlation matrix $\mathbf{C}$,

$$\mathbf{C} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T. \quad (18)$$

Here $\Lambda \equiv \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N$ are the eigenvalues in nonincreasing order, $\mathbf{U}$ is an orthonormal matrix, whose ith column is the basis eigenvector $\mathbf{u}_i$ of $\mathbf{C}$, and $\mathbf{U}^T$ is the transpose of $\mathbf{U}$, which is equal to $\mathbf{U}^{-1}$ because of orthonormality.

For each eigenvalue and the corresponding eigenvector, it holds that

$$\lambda_i = \mathbf{u}_i^T\mathbf{C}\mathbf{u}_i = \mathbf{u}_i^T\mathrm{Cov}(\mathbf{z}_t)\mathbf{u}_i = \mathrm{Var}\big(\mathbf{u}_i^T\mathbf{z}_t\big) = \mathrm{Var}(\alpha_{i,t}). \quad (19)$$

According to PCA, $\alpha_{i,t} = \mathbf{u}_i^T\mathbf{z}_t$ is defined as the ith principal component ($\alpha_{i,t}$), and the eigenvalue $\lambda_i = \mathrm{Var}(z_{i,t})$ indicates the portion of total variance of $\mathbf{z}_t$ contributed to $\alpha_{i,t}$, as shown in Eq. (19). Since the total variance of $\mathbf{z}_t$ is

$$\sum_{i=1}^{N}\mathrm{Var}(z_{i,t}) = \mathrm{tr}(\mathbf{C}) = \sum_{i=1}^{N}\lambda_i, \quad (20)$$

the expression $\lambda_i/\mathrm{tr}(\mathbf{C})$ indicates the percentage of the total variance of $\mathbf{z}_t$ that can be explained by the $\alpha_{i,t}$. According to PCA (a) the principal components $\alpha_{i,t}$ are uncorrelated with each other and (b) $\alpha_{i,t}$ maximizes the variance of the linear combination of $\mathbf{U}^T\mathbf{z}_t$ with the orthonormal restriction $\mathbf{U}^T\mathbf{U} = 1$ given the previous principal components [27].

From the orthonormal property of $\mathbf{U}$ we obtain

$$\mathbf{I} = \mathbf{U}\mathbf{U}^T = \mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \cdots + \mathbf{u}_N\mathbf{u}_N^T, \quad (21)$$

where $\mathbf{I}$ is the identity matrix. Then the multiple time series $\mathbf{z}_t$ can be represented as a linear combination of all the $\boldsymbol{\alpha}_t$

$$\begin{aligned}\mathbf{z}_t &= \big(\mathbf{u}_1\mathbf{u}_1^T + \mathbf{u}_2\mathbf{u}_2^T + \cdots + \mathbf{u}_N\mathbf{u}_N^T\big)\mathbf{z}_t \\ &= \mathbf{u}_1\alpha_{1,t} + \mathbf{u}_2\alpha_{2,t} + \cdots + \mathbf{u}_N\alpha_{N,t}. \end{aligned} \quad (22)$$

The total variance of all time series can be proved to be equal to the total variance of all principal components

$$\sum_{i=1}^{N}\mathrm{Var}(z_{i,t}) = \mathrm{Var}(\mathbf{u}_1)\alpha_{1,t} + \cdots + \mathrm{Var}(\mathbf{u}_N)\alpha_{N,t} \quad (23)$$

$$= \sum_{i=1}^{N}\mathbf{u}_i^T\mathbf{u}_i\,\mathrm{Var}(\alpha_{i,t}) = \sum_{i=1}^{N}\mathrm{Var}(\alpha_{i,t}). \quad (24)$$

Next we assume that $\mathrm{Var}(\alpha_{1,t}) = \lambda_1$ is much larger than each of the rest of the eigenvalues, which means that the first $\boldsymbol{\alpha}_t$, $\alpha_{1,t}$, accounts for most of the total variances of all the time series. We express $\mathbf{z}_t$ as the sum of the first part of Eq. (22) corresponding to $\alpha_{1,t}$ and the error term combined from all other terms in Eq. (22). Thus,

$$\mathbf{z}_t = \mathbf{u}_1\alpha_{1,t} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \equiv \sum_{i=2}^{N}\mathbf{u}_i\alpha_{i,t}. \quad (25)$$

Then $\alpha_{1,t}$ is a good approximation of the global factor M_t because it is a linear combination of $R_{i,t}$ that accounts for the largest amount of the variance. α_1 is a zero-mean process because it is a linear combination of $z_{i,t}$ which are also zero-mean processes [see Eq. (17)].

Comparing Eqs. (17) and (25) with

$$R_{i,t} = \mu_i + b_iM_t + \epsilon_{i,t}, \quad (26)$$

we find the following estimates:

$$\begin{aligned} M_t &= \alpha_{1,t}, \\ b_i &= \sigma(R_i)u_{1i}, \\ \epsilon_{i,t} &= \sigma(R_i)\eta_{i,t}. \end{aligned} \quad (27)$$

Using Eq. (19) we find that the cross correlation between $|M_t|$ and $|R_{i,t}|$

$$\mathrm{Corr}(M_t, R_{i,t}) = \sqrt{\lambda_i}u_{i1}. \quad (28)$$

In the rest of this work, we apply the method of Eq. (27) to empirical data.

B. Analysis of the global factor

Next we apply the method of Eq. (27) to estimate the global factor of 48 world index returns. We calculate the autocorrelations of M_t and $|M_t|$, which are shown in Figs. 3 and 4. Precisely, for the world indices, Fig. 3(a) shows the time series of the global factor M_t, and Fig. 3(b) shows the autocorrelations in M_t. We find only short-range autocorrelations because, after an interval $\Delta t = 2$, most autocorrelations in M_t fall in the range of $(-1.96\sqrt{1/T}, 1.96\sqrt{1/T})$ [61], which is the 95% confidence interval for zero autocorrelations, Here $T = 2744$.

For the 48 world index returns, Fig. 4(a) shows the time series of magnitudes $|M_t|$, with few clusters related

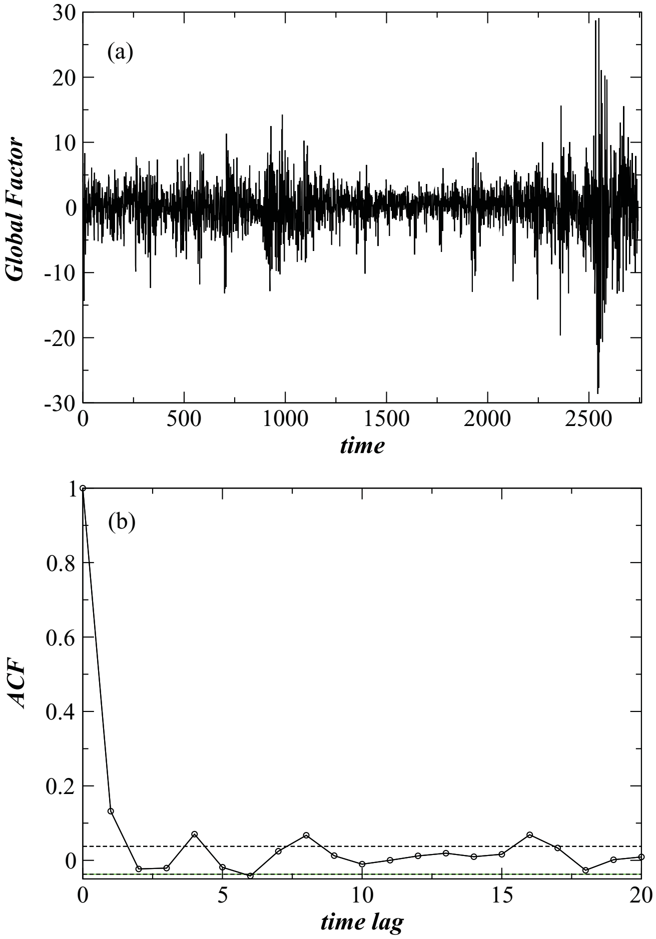

FIG. 3. Short-range autocorrelations of the global factor. (a) Time series of the global factor. (b) The autocorrelation function (ACF) of the global factor. The region between dashed lines is the 95% confidence interval for the no autocorrelation hypothesis. Autocorrelations are smaller than 0.132 except $\Delta t = 0$, and become insignificant after time lag $\Delta t = 2$, with no more than one significant autocorrelation for every 20 time lags. Therefore, only short-range autocorrelations can be found in the global factor.

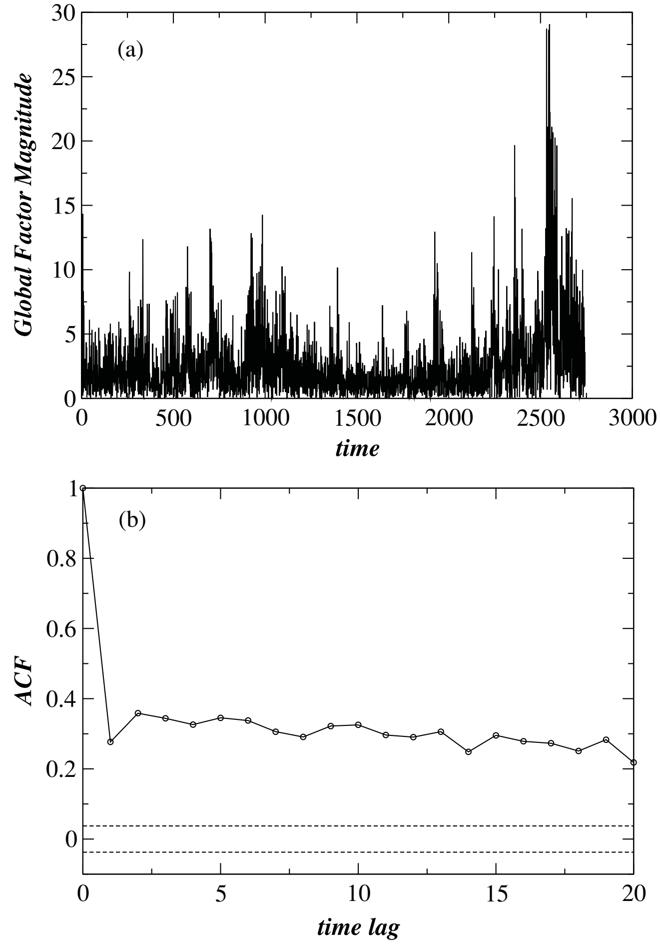

FIG. 4. Long-range autocorrelations of the magnitude global factor. (a) Time series of magnitudes of the global factor. (b) Autocorrelations of magnitudes of the global factor. The region between dashed lines is the 95% confidence interval for the no autocorrelation hypothesis. Autocorrelations are much larger than the autocorrelations of the global factor itself, as large as 0.359 at $\Delta t = 2$, and is still larger than 0.2 until $\Delta t = 33$. For every time lag, the autocorrelation is significant even after $\Delta t = 100$. Therefore long-range autocorrelations exist in the magnitudes of the global factor.

to market shocks during which the market fluctuates more. Figure 4(b) shows that, in contrast to M_t, the magnitudes $|M_t|$ exhibit long-range autocorrelations since the values $|M_t|$ are significant even after $\Delta t = 100$. The autocorrelation properties of the global factor are the same as the autocorrelation properties of the individual indices (i.e., there are short-range autocorrelations in M_t and long-range power-law autocorrelations in $|M_t|$ [33,62]). These results are also in agreement with Fig. 1(b) where the largest singular value λ_L vs Δt calculated for $\tilde{\mathsf{C}}$ decays more slowly than the largest singular value calculated for C. As found in Ref. [26] for $\Delta t \gg 1, \lambda_L(\Delta t)$ approximately follows the same decay pattern as cross correlation functions. Although a Ljung-Box test shows that the return autocorrelation is significant for a 95% confidence level [65], the return autocorrelation is only 0.132 for $\Delta t = 1$ and becomes insignificant after $\Delta t = 2$. Therefore, for simplicity, we only consider magnitude cross correlations in modeling the global factor.

We model the long-range global factor **M** with a particular version of the GARCH process, the Glosten-Jagannathan-Runkle (GJR)GARCH process [66], because this GARCH version explains well the asymmetry in volatilities found in many world indices [66–68]. The GJR GARCH model can be written as

$$\epsilon_t = \sigma_t \eta_t, \tag{29}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} (\alpha_i + \gamma T_{t-i}) \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2, \tag{30}$$

where σ_t is the volatility and η_t is a random process with a Gaussian distribution with standard deviation 1 and mean 0. The coefficients α and β are determined by a maximum likelihood estimation (MLE) and $T_t = 1$ if $\epsilon_{t-1} < 0$, $T_t = 0$ if $\epsilon_{t-1} \geqslant 0$. We expect the parameter γ to be positive, implying that bad news (negative increments) increases volatility more

TABLE I. GJR GARCH(1,1) coefficients of the global factor. The P values and t values confirm that all these parameters are significant at 95% confidence level. The positive value of γ means bad news has larger impact on the global market than good news. We find $\alpha_1 + \beta_1 + \gamma/2 = 0.9756$, which is very close to one, and so indicates long-range volatility autocorrelations.

	Value	Standard Error	t value	P value
α_0	0.2486	0.0283	8.789	0.0000
α_1	0.0170	0.0080	2.128	0.0334
β_1	0.8790	0.0101	86.939	0.0000
γ	0.1591	0.0148	10.805	0.0000

than good news. For the sake of simplicity, we follow the usual procedure of setting $p = q = 1$ in all numerical simulations. In this case, the GJR GARCH(1,1) model for the global factor can be written as

$$M_t = \sigma_t \eta_t, \quad (31)$$

$$\sigma_t^2 = \alpha_0 + (\alpha_1 + \gamma T_{t-1})\epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (32)$$

We estimate the coefficients in the above equations using MLE, where the estimated coefficients are shown in Table I.

Next we test the hypothesis that a significant percentage of the world cross correlations can be explained by the global factor. By using PCA we find that the global factor can account for 30.75% of the total variance. Note that, according to RMT, only the eigenvalues larger than the largest eigenvalue of a Wishart matrix calculated by Eq. (3) (and the corresponding $\boldsymbol{\alpha}$'s) are significant. To calculate the percentage of variance accounted for by the significant $\boldsymbol{\alpha}$'s, we employ the RMT approach proposed in Ref. [15]. The largest eigenvalue for a Wishart matrix is $\lambda_+ = 1.282$ for $N = 48$ and $T = 2744$ as found in the empirical data. From all the 48 eigenvalues, only the first three are significant: $\lambda_1 = 14.762$, $\lambda_2 = 3.453$, and $\lambda_3 = 1.380$. This result implies that among the significant factors, the global factor accounts for $\lambda_1 / \sum_{i=1}^{3} \lambda_i = 75.34\%$ of the variance, confirming our hypothesis that the global factor accounts for most of the variance of all individual index returns.

PCA is defined to estimate the percentage of variance the global factor can account for in zero-time-lag correlations.Next we study the time-lag cross correlations after removing the global trend, and apply the SVD to the correlation matrix of regression residuals η_i of each index [see Eq. (8)]. Our results show that for both returns and magnitudes, the remaining cross correlations are very small for all time lags compared to cross correlations obtained for the original time series. This result additionally confirms that a large fraction of the world cross correlations for both returns and magnitudes can be explained by the global factor.

VI. APPLICATIONS OF GLOBAL FACTOR MODEL

A. Locating and forecasting global risks

The asymptotic (unconditional) variance for the GJR GARCH model is $\alpha_0/(1 - \alpha_1 - \beta_1 - \gamma/2) = 10.190$ [69]. For the global factor, the conditional volatility σ_t can be estimated by recursion using the historical conditional volatilities

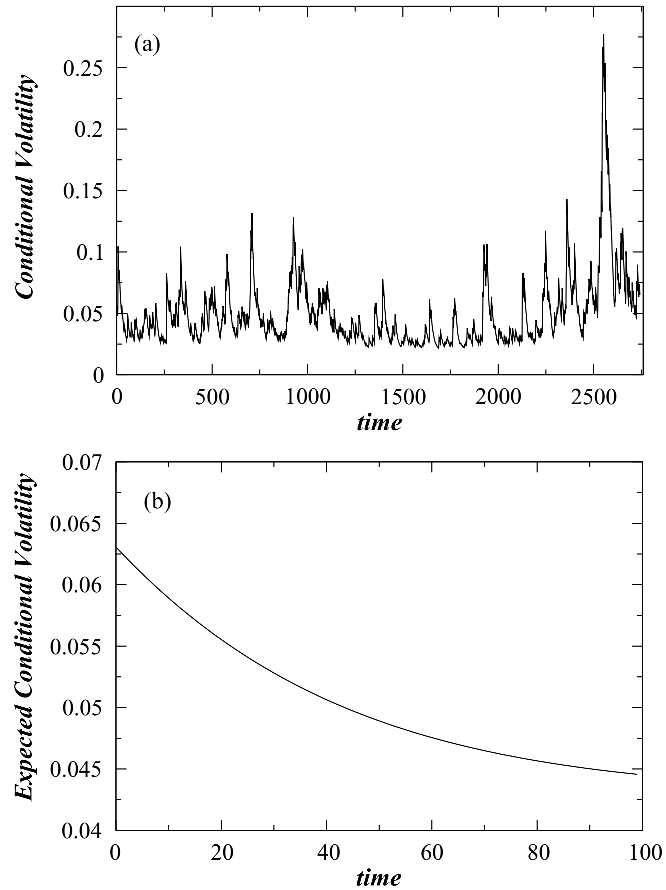

FIG. 5. (a) Conditional volatility of the global factor, showing that the clusters in the conditional volatilities may serve to predict market crashes. In each cluster, the height indicates the size of the market crash, and the width indicates its duration. (b) The 100-day forecasted volatility of the global factor, using the past data ranging from 4 Jan 1999 through 10 July 2009. It will converge to the unconditional volatility asymptotically.

and fitted coefficients in Eq. (32). For example, the largest cluster at the end of the graph shows the 2008 financial crisis. In Fig. 5(a) we show the time series of the conditional volatility of Eq. (32) of the global factor. The clusters in the conditional volatilities may serve to predict market crashes. In each cluster, the height is a measure of the size of the market crash, and the width indicates its duration. In Fig. 5(b) we show the forecasting of the conditional volatility of the global factor, which asymptotically converges to the unconditional volatility.

B. Finding uncorrelated individual indices

Next, in Fig. 6 we show the cross correlations between the global factor and each individual index using Eq. (28). There are indices for which cross correlations with the global factor are very small compared to the other indices; 10 of 48 indices have cross correlation coefficients with the global factor smaller than 0.1. These indices correspond to Iceland, Malta, Nigeria, Kenya, Israel, Oman, Qatar, Pakistan, Sri Lanka, and Mongolia. The financial market of each of these countries is weakly bound with financial markets of other countries. This is useful information for investment managers

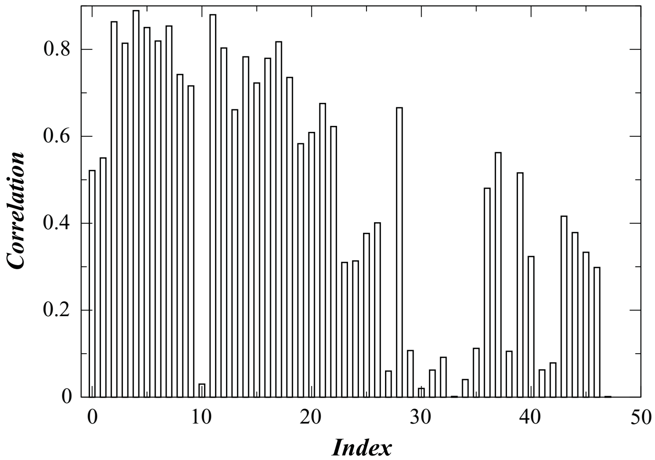

FIG. 6. Cross correlation between the global factor M_t and each individual index $R_{i,t}$, $i = 1,2,\ldots,48$. The global factor has large correlation with most of the indices. However, there are indices that are not much correlated with the global factor. Of the 48 indices, ten have a correlation smaller than 0.1 between the global factor, corresponding to the indices for Iceland, Malta, Nigeria, Kenya, Israel, Oman, Qatar, Pakistan, Sri Lanka, and Mongolia. Hence, unlike most countries, the economies of these ten countries are more independent of the world economy.

because one can reduce risk by investing in these countries during world market crashes which seemingly do not influence these countries severely.

VII. DISCUSSION

We have developed a modified time-lag random matrix theory in order to quantify the time-lag cross correlations among multiple time series. Applying the modified TLRMT to the daily data for 48 worldwide financial indices, we find short-range cross correlations between the returns and long-range cross correlations between their magnitudes. The magnitude cross correlations show a power-law decay with time lag, and the scaling exponent is 0.25. The result we obtain, that at the world level the cross correlations between the magnitudes are long range, is potentially significant because it implies that strong market crashes introduced at one place have an extended duration elsewhere, which is bad news for international investment managers who imagine that diversification across countries reduces risk.

We model long-range world index cross correlations by introducing a global factor model in which the time-lag cross correlations between returns (magnitudes) can be explained by the autocorrelations of the returns (magnitudes) of the global factor. We estimate the global factor as the first component by using principal component analysis. Using random matrix theory, we find that only three principal components are significant in explaining the cross correlations. The global factor accounts for 30.75% of the total variance of all index returns, and 75.34% of the variance of the three significant principal components. Therefore, in most cases, a single global factor is sufficient.

We also show the applications of the GFM, including locating and forecasting world risk, and finding individual indices that are weakly correlated to the world economy. Locating and forecasting world risk can be realized by fitting the global factor using a GJR GARCH(1,1) model, which explains both the volatility correlations and the asymmetry in the volatility response to both good news and bad news. The conditional volatilities calculated after fitting the GJR GARCH(1,1) model indicate the global risk, and the risk can be forecast by recursion using the historical conditional volatilities and the fitted coefficients. To find the indices that are weakly correlated to the world economy, we calculate the correlation between the global factor and each individual index. We find ten indices which have a correlation smaller than 0.1, while most indices are strongly correlated to the global factor with correlations larger than 0.3. To reduce risk, investment managers can increase the proportion of investment in these countries during world market crashes, which do not severely influence these countries.

Based on principal component analysis, we propose a general method which helps extract the most significant components in explaining long-range cross correlations. This makes the method suitable for a broad range of phenomena where time series are measured, ranging from seismology and physiology to atmospheric geophysics. We expect that the cross correlations in EEG signals are dominated by the small number of most significant components controlling the cross correlations. We speculate that cross correlations in earthquake data are also controlled by some major components. Thus the method may have significant predictive and diagnostic power that could prove useful in a wide range of scientific fields.

ACKNOWLEDGMENTS

We thank Ivo Grosse and T. Preis for valuable discussions and NSF for financial support.

[1] *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers (Springer, Berlin, 2009).

[2] K. Christensen and N. R. Moloney, *Complexity and Criticality* (Imperial College Press, London, 2010).

[3] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, 2010).

[4] Heartbeat interval time series are among many time series characterizing the functioning human.

[5] P. Samuelsson, E. V. Sukhorukov, and M. Buttiker, Phys. Rev. Lett. **91**, 157002 (2003).

[6] A. Cottet, W. Belzig, and C. Bruder, Phys. Rev. Lett. **92**, 206801 (2004).

[7] I. Neder, M. Heiblum, D. Mahalu, and V. Umansky, Phys. Rev. Lett. **98**, 036803 (2007).

[8] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).

[9] A. Corral, Phys. Rev. Lett. **95**, 159801 (2005).

[10] M. Campillo and A. Paul, Science **299**, 547 (2003).
[11] E. Lippiello, L. de Arcangelis, and C. Godano, Phys. Rev. Lett. **100**, 038501 (2008).
[12] B. Solnik, C. Bourcrelle, and Y. Le Fur, Financ. Anal. J. **52**, 17 (1996).
[13] C. B. Erb, C. R. Harvey, and T. E. Viscanta, Financ. Anal. J. **50**, 32 (1994).
[14] B. LeBaron, W. B. Arthur, and R. Palmer, J. Econ. Dyn. Control **23**, 1487 (1999).
[15] L. Laloux, P. Cizeau, J. P. Bouchaud, M. Potters, Phys. Rev. Lett. **83**, 1467 (1999); V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley, *ibid.* **83**, 1471 (1999); V. Plerou P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, H. E. Stanley, Phys. Rev. E **65**, 066126 (2002).
[16] M. Takayasu and H. Takayasu, *Statistical Physics and Economics* (Cambridge University Press, Cambridge, 2010).
[17] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).
[18] L. Kullmann, J. Kertesz, and K. Kaski, Phys. Rev. E **66**, 026125 (2002).
[19] M. Tumminello, T. Aste, T. Di Matteo, and R. N. Mantegna, Proc. Natl. Acad. Sci. USA **102**, 10421 (2005).
[20] T. Mizuno, H. Takayasu, and M. Takayasu, Physica A **364**, 336 (2006).
[21] B. Podobnik and H. E. Stanley, Phys. Rev. Lett. **100**, 084102 (2008); B. Podobnik *et al.*, Proc. Natl. Acad. Sci. USA **106**, 22079 (2009).
[22] S. Arianos and A. Carbone, J. Stat. Mech. (2009) P03037.
[23] B. Podobnik *et al.*, Eur. Phys. J. B **56**, 47 (2007).
[24] H. M. Markowitz, J. Financ. **7**, 77 (1952).
[25] P. Bob, M. Susta, K. Glaslova, and N. N. Boutros, Psych. Res. **177**, 37 (2010).
[26] B. Podobnik *et al.*, Europhys. Lett. **90**, 68001 (2010).
[27] I. T. Jolliffe, *Principal Component Analysis* (Springer-Verlag, New York, 1986).
[28] T. Guhr and B. Kalber, J. Phys. A **36**, 3009 (2003).
[29] C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994); A. Carbone, G. Castelli, and H. E. Stanley, Physica A **344**, 267 (2004); L. Xu, C. P. Ivanov, K. Hu, Z. Chen, A. Carbone, H. E. Stanley, Phys. Rev. E **71**, 051101 (2005); A. Carbone, G. Castelli, and H. E. Stanley, *ibid.* **69**, 026105 (2004).
[30] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden Day, San Francisco, 1970).
[31] R. F. Engle, Econometrica **50**, 987 (1982).
[32] T. Bollerslev, J. Econometr. **31**, 307 (1986).
[33] Z. Ding and C. W. J. Granger, J. Econometr. **73**, 185 (1996).
[34] Y. Ashkenazy, P. C. Ivanov, S. Havlin, C.-K. Peng, A. L. Goldberger, H. E. Stanley, Phys. Rev. Lett. **86**, 1900 (2001).
[35] J. W. Kantelhard, Y. Ashkenazy, P. C. Ivanov,A. Bunde, S. Havlin, T. Penzel, J. H. Peter, H. E. Stanley, Phys. Rev. E **65**, 051908 (2002).
[36] V. N. Livina, Y. Ashkenazy, P. Braun, R. Monetti, A. Bunde, S. Havlin, Phys. Rev. E **67**, 042101 (2003).
[37] B. Podobnik, P. C. Ivanov, V. Jazbinsek, Ƶ. Trontelj, H. E. Stanley, and I. Grosse, Phys. Rev. E **71**, 025104(R) (2005).
[38] W. F. Sharpe, J. Financ. **19**, 425 (1964).
[39] W. F. Sharpe, *Portfolio Theory and Capital Markets* (McGraw-Hill, New York, 1970).
[40] T. Bollerslev, R. F. Engle, and J. M. A. Wooldridge, Rev. Econ. Stat. **72**, 498 (1988).
[41] T. Bollerslev, R. Engle, and J. Wooldridge, Journal of Political Economy **96**, 116 (1988).
[42] R. Engle and F. K. Kroner, Econometr. Theory **11**, 122 (1995).
[43] G. Canarella and S. K. Pollard, Int. Rev. Econ. **54**, 445 (2007).
[44] Z. Ding, C. W. J. Granger, and R. F. Engle, J. Empirical Financ. **1**, 83 (1993).
[45] FTSE 100, DAX, CAC 40, IBEX 35, Swiss Market, FTSE MIB, PSI 20, Irish overall, OMX Iceland 15, AEX, BEL 20, Luxembourg LuxX, OMX Copenhagen 20, OMX Helsinki, OBX Stock, OMX Stockholm 30, Austrian Traded ATX, Athex Composite Share Price, WSE WIG, Prague Stock Exch, Budapest Stock Exch INDX, Bucharest BET Index, SBI20 Slovenian Total Mt, OMX Tallin OMXT, Malta Stock Exchange IND, FTSE/JSE Africa TOP40 IX.
[46] ISE National 100, Tel Aviv 25, msm30, dsm 20, Mauritius Stock Exchange, NIKKEI 225, Hang Seng, Shanghai se b share, all Ordinaries, nzx all, Karachi 100, Sri Lanka Colombo all sh, Stock Exch Of Thai, Jakarta Composite, FTSE Bursa Malaysia KLCI, PSEi - Philippine SE, MSE Top 20.
[47] S&P 500, Mexico BOLSA.
[48] FTSE/JSE Africa TOP40 IX, CFG 25, Nigerian Stock Exchange All Shares, Nairobi Stock Exchange 20-Share.
[49] F. Longin and B. Solnik, J. Int. Money Financ. **14**, 3 (1995).
[50] L. Ramchand and R. Susmel, J. Empirical Financ. **5**, 397 (1998).
[51] M. L. Mehta, *Random Matrices* (Academic Press, Boston, 1991).
[52] T. Guhr, A. Müller-Groeling, and H. Weidenmüller, Phys. Rep. **299**, 190 (1998).
[53] L. Laloux, P. Cizeau, J. P. Bouchaud, and M. Potters, Risk Magazine **12**, 69 (1999).
[54] A. Utsugi, K. Ino, and M. Oshikawa, Phys. Rev. E **70**, 026110 (2004).
[55] R. K. Pan and S. Sinha, Phys. Rev. E **76**, 046116 (2007).
[56] J. Shen and B. Zheng, Europhys. Lett. **86**, 48005 (2009); T. Conlon, H. J. Ruskin, M. Crane, Physica A **388**, 705 (2009).
[57] M. Potters, J.-P. Bouchaud, L. Laloux, Acta Phys. Pol. B **36**, 2767 (2005).
[58] J. Kwapien, S. Drozdz, A. Z. Gorski, P. Oswiecimka, Acta Phys. Pol. B **37**, 3039 (2006).
[59] J.-P. Bouchaud *et al.*, Eur. Phys. J. B **55**, 201 (2007).
[60] A. M. Sengupta and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[61] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications* (Springer-Verlag, New York, 2000).
[62] Y. Liu *et al.*, Physica A **245**, 437 (1997); Y. Liu, P. Gopikrishna, P. Cizeau, M. Meyer, C. K. Peng, H. E. Stanley, Phys. Rev. E **60**, 1390 (1999); P. Cizeau *et al.*, Physica A **245**, 441 (1997).
[63] S. A. Ross, J. Econ. Theory **13**, 341 (1976).
[64] W. F. Sharpe, Manage. Sci. **9**, 277 (1963).
[65] G. M. Ljung and G. E. P. Box, Biometrika **65**, 297 (1978).
[66] L. Glosten, R. Jagannathan, and D. Runkle, J. Financ. **48**, 1779 (1993).
[67] S. A. Jayasuriya and R. Rossiter, J. Int. Financ. Econ. **8**, 11 (2008).
[68] J. Tenenbaum D. Horvatic, S. C. Bajic, B. Pehlivanovic, B. Podobnik, and H. E. Stanley, Phys. Rev. E **82**, 046104 (2010).
[69] S. Ling and M. McAleer, J. Econometr. **106**, 109 (2002).