# Phase distribution and phase correlation of financial time series

Ming-Chya Wu,[1,*] Ming-Chang Huang,[2] Hai-Chin Yu,[3] and Thomas C. Chiang[4]
[1]*Institute of Physics, Academia Sinica, Nankang, Taipei 11529, Taiwan*
[2]*Department of Physics, Chung-Yuan University, Chungli, Taiwan 32023*
[3]*Department of International Business, Chung-Yuan University, Chungli, Taiwan 32023*
[4]*Department of Finance, Drexel University, Philadelphia, Pennsylvania 19104, USA*
(Received 22 July 2005; published 13 January 2006; publisher error corrected 18 January 2006)

The scaling, phase distribution, and phase correlation of financial time series are investigated based on the Dow Jones Industry Average and NASDAQ 10-min intraday data for a period from 1 Aug. 1997 to 31 Dec. 2003. The returns of the two indices are shown to have nice scaling behaviors and belong to stable distributions according to the criterion of Lévy's $\alpha$ stable distribution condition. An approach catching characteristic features of financial time series based on the concept of instantaneous phase is further proposed to study the phase distribution and correlation. Analysis of the phase distribution concludes that return time series fall into a class which is different from other nonstationary time series. The correlation between returns of the two indices probed by the distribution of phase difference indicates that there was a remarkable change of trading activities after the event of the 9/11 attack, and this change persisted in later trading activities.

DOI: 10.1103/PhysRevE.73.016118 PACS number(s): 89.65.Gh, 05.45.Tp, 05.40.Fb

## I. INTRODUCTION

Financial markets are complex systems consisting of a large number of traders interacting with one another in the market and reacting to external information to determine the best price for a given item. Previous studies of financial systems are usually based on fundamental statistics on return of index and tend to address issues on drawing trading strategies for traders and investors. With the power of new algorithms for statistical analysis, some previous studies have provided rich information for such purposes. For example, many existing works in the literature disclose that high-frequency data of financial assets would show a heavy tail. Zhou indicated that the heavy tail of financial time series is mostly caused by the heteroscedasticity of the time series [1]. Furthermore, the estimates of intraday volatility basically reveal a strong seasonal pattern.

However, previous studies have also suffered from a limited scope provided by the statistics of conventional derivatives from returns. As a result, cross-disciplinary studies on financial systems have attracted much attention in recent decades [2–7]. With the help of ideas from other fields, there have been significant achievements in economy science. One of great achievements has been applications of statistical mechanics to economic systems, which was later referred to *econophysics* [2]. Some correspondences between quantities in economic systems and physical systems were found and suggest fundamental concepts behind them. For example, it was found that there is a two-phase behavior of financial markets which suggests that there is a link between the dynamics of a human system with many interacting participants and the ubiquitous phenomenon of phase transitions that occur in physical systems with many interacting units [8]. The scaling analysis in statistical mechanics is shown to be applicable in the studies of market systems [2,3,9].

There are also developments in the analysis of financial time series in methodology [4,5,10]. For example, the method of random matrix theory has been developed to study the statistical structure of multivariate time series [4–6] and has given remarkable agreement between theoretical prediction and empirical data concerning the density of eigenvalues associated with the time series of the different stocks of the S&P500 [4,6]. Furthermore, the wavelet transform modulus maximum approach has been available in the last ten years [11] and has been applied to study nonstationary time series such as physiologic systems [12–15] and economic systems [10]. For example, Ohashi *et al.* [10] used the analysis of asymmetrical singularities to analyze the human heartbeat and daily stock price records, and claimed that the method can enhance understanding of the mechanisms determining the systems' dynamics [10]. However, the wavelet analysis has difficulty because of its nonadaptive nature, so that once the basic wavelet is selected, it is used to analyze all the data. In addition, some wavelets are Fourier based, suffering the shortcoming of Fourier spectral analysis for only giving a physical meaningful interpretation to linear phenomena [16]. Nevertheless, financial time series are nonlinear, and the analyses by these approach may lose information on nonlinear properties.

In this paper, in light of the above situations, we intend to develop an approach to study financial time series. The main issues to be addressed are scaling analysis and the phase distribution of financial time series and phase correlation between them. We use the Dow Jones Industrial Average 30 (DJIA) and NASDAQ stock indices for the study. The reason to choose these two indices stems from the fact that the former represents the most established and renowned firms in the U.S. market, while the latter consists of high-tech and growth firms. These two indices thus represent not only the core of the U.S. economy, but also facilitate the menu for investor's choice in the mean-variance plan. A successful empirical investigation emerging from this study is bound to

*Electronic address: mcwu@phys.sinica.edu.tw

 

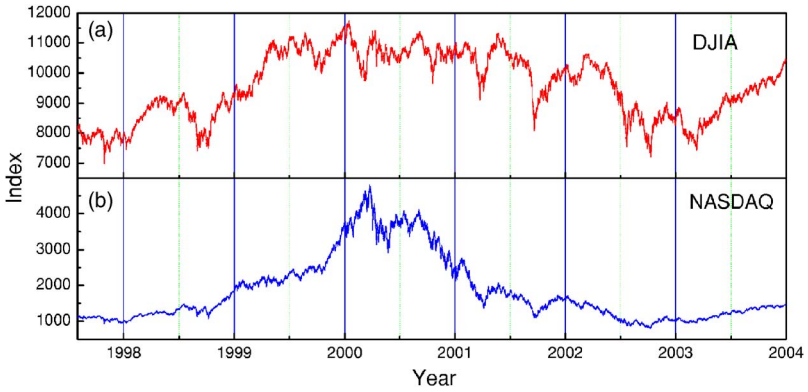


FIG. 1. (Color online) High-frequency (a) DJIA and (b) NASDAQ index date sampled by 10 min from 1 Aug. 1997 to 31 Dec. 2003.

provide a solid foundation to advance our empirical analyses to a broader spectrum of asset behavior.

We first analyze scaling behaviors of return time series to probe stability of their distributions. Following the scheme proposed in Refs. [2,9], scaling analyses are performed on the returns of the DJIA and NASDAQ with various time sampling intervals. The stock returns of both the DJIA and NASDAQ are shown to have nice scaling behaviors and belong to stable distributions according to the criterion of Lévy's $\alpha$ stable distribution condition [3].

We further suggest an approach to examine properties of financial time series. The Hilbert-Huang time signal analysis method [16] is used to define and evaluate instantaneous phase of financial time series. This method was first disclosed by Huang *et al.* for the studies of nonstationary and nonlinear time series [16]. The key part of the method is the empirical mode decomposition (EMD) method with which any complicated data set can be decomposed into a finite and often small number of intrinsic mode functions (IMF's) that admit well-behaved Hilbert transforms [16]. Here, based on this method we analyze the nonstationary time series of return to extract characteristic structures of empirical data and evaluate phase distributions of certain IMF's. Then, the resultant structures can be further analyzed. The results we thus obtain indicate that the return time series fall into a class which is different from other time series.

We also measure the correlation between DJIA and NASDAQ indices by calculating the distribution of phase difference in return time series. Our results impressively show that there is a remarkable change of trading activities implied by phase correlation after the event of the 9/11 attack. For modern stock markets with information transmitted rapidly, our findings might be a useful reference for market investors and policy decision makers [17].

This paper is organized as follows. In Sec. II, we briefly illustrate the data source used in this paper. In Sec. III, we give basic definitions of quantities used in sequential analyses and explore general features presented by these quantities. Then, we start to perform scaling analysis of returns in Sec. IV. In Sec. V, we suggest our approach for catching characteristic features of financial time series. The Hilbert-Huang method is used to decompose return time series and calculate instantaneous phases accordingly. Correlations between two financial time series are investigated and discussed in Sec. VI. Finally, we summarize our results in Sec. VII.

## II. DATA

The empirical analyses are based on DJIA and NASDAQ from the Trade And Quotation (TAQ) database and the Yahoo database [18]. The TAQ data files contain continuously recorded information on the trades and quotations for the securities listed on the NYSE, AMEX, and NASDAQ. The DJIA stocks are the most actively traded securities; the capital size of the firms in the DJIA also helps to ensure a high degree of liquidity. Alternatively, the stocks listing in NASDAQ exchanges characterize high-tech growing firms, yet with more price volatility. On the basis of these two distinguished characteristics of stock returns, we are able to derive some empirical regularity on diverse portfolios.

The intraday 10-min scale values for both the DJIA and NASDAQ spanning from 1 Aug. 1997 through 31 Dec. 2003 cover the whole six-and-half hours trading starting from 9:30 to 15:50 EST. The overnight (or over weekend) period constitutes an unusual time period as it involves a much longer time interval than 10 min. Therefore, the value of index at very open price will be distorted. Here, unlike Main and Adam [19], we do not omit the very close-to-open returns. Rather, we keep them in the data to conduct sensitivity analysis. After elimination of the omitting days for which all the 10-min values of the index were not available, we obtain a total of 1543 trading days with 60 177 observations of 10-min index values. Following the analyses by Andersen and Bollerslev [20], we constructed 10-min returns with the daily transaction records extending from 9:30 to 15:50 EST, a total of 39 10-min returns for each day. The 10-min horizon is short enough that the accuracy of the continuous records of realized returns and volatility work well, and it is long enough that the confounding influences from market microstructure frictions can be negligible.

Figures 1(a) and 1(b) are time series paths of the DJIA and NASDAQ indices sampled by 10 min.

## III. TIME SERIES OF RETURNS

We first define basic quantities used in this paper and present general features revealed from the data. The index values are denoted by a time series $Y(t)$, and the time series of logarithmic returns of an asset priced at $Y(t)$ over a time scale $\tau$ is defined as

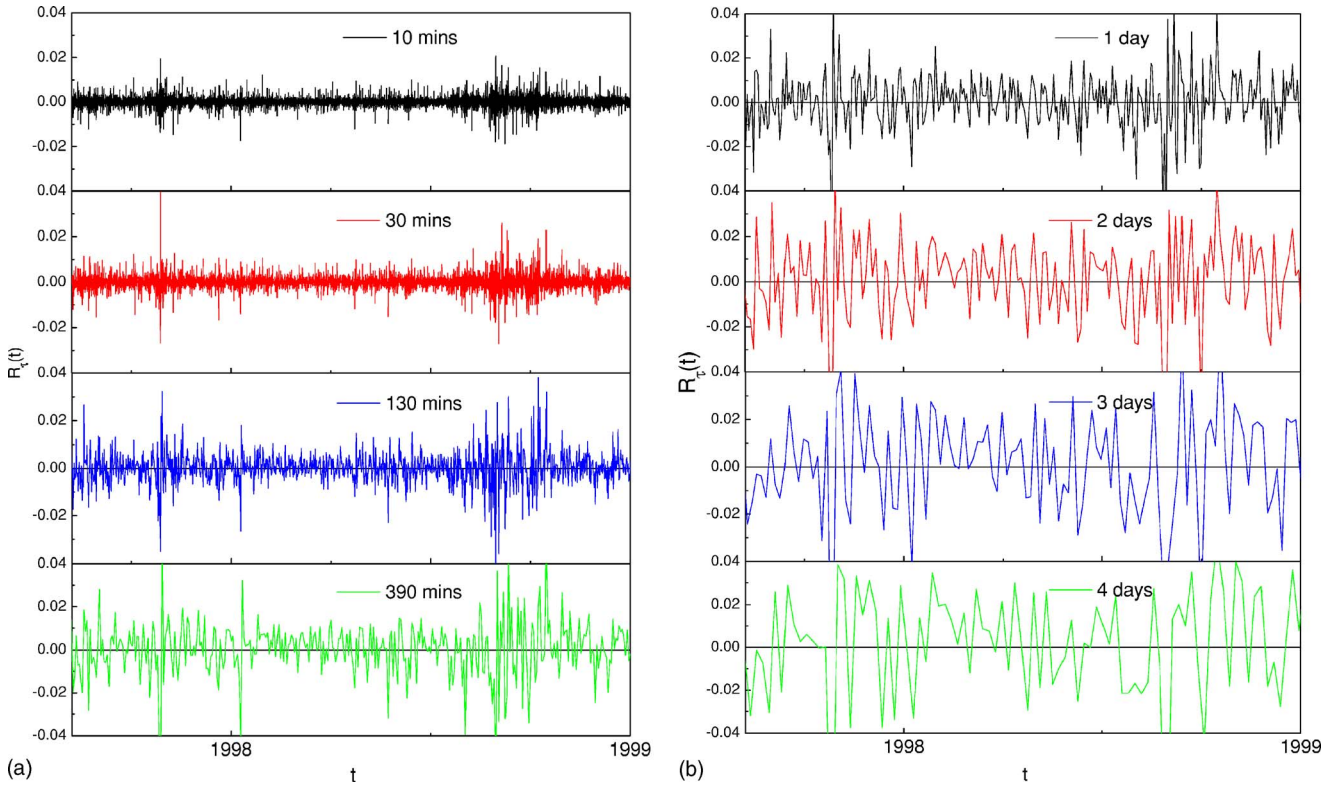


FIG. 2. (Color online) Time series of logarithmic returns $R_\tau(t)$ of (a) the intraday DJIA index sampled by 10, 30, 130, and 390 min and (b) the interday DJIA index sampled by 1, 2, 3, and 4 days.

$$R_\tau(t) = \ln\left[\frac{Y(t)}{Y(t-\tau)}\right], \tag{1}$$

where $\tau$ is a multiple of the primary time sampling unit $\Delta t$(=10 min). Since the time scale $\tau$ (in unit of $\Delta t$) is a parameter used to sample time series of returns, we can take different $\tau$ for $R_\tau(t)$ to explore the behaviors of the returns with intraday and interday frequencies. Because there are 39 sampling data in each trading day, we take different $\tau$ values being factors of 390 to avoid ambiguity in time sampling intervals involving interday quantities. Here we take $\tau = 10, 30, 130, 390$ min to sample the time series of $R_\tau(t)$ for the intraday data and $\tau = 1, 2, 3, 4$ days for daily data, and the results for DJIA are shown in Figs. 2(a) and 2(b), respectively. These figures essentially provide a picture that amplitudes of these time series are, in general, proportional to sample time scales. Therefore, based on Eq. (1), we define the normalized logarithmic returns as [3]

$$r_\tau(t) = \frac{R_\tau(t) - \langle R_\tau(t)\rangle}{\sqrt{\langle R_\tau^2(t)\rangle - \langle R_\tau(t)\rangle^2}}, \tag{2}$$

where the expectation values denoted by $\langle\cdot\rangle$ are taken over the entire time period under consideration.

Be aware of the fact that the analysis based on the assumption that the tick-by-tick data are linear may lead to incorrect conclusions if the underlying process of financial time series is multiplicative [21]. We will use the returns defined in Eq. (1) instead of the index changes $Y(t) - Y(t-\tau)$, used in Ref. [9], for the scaling analysis in the next section.

We further define the probability distribution (or more precisely, probability density function) $P$ as the normalized distribution of a measure $\rho$, which satisfies

$$\int_{-\infty}^{\infty} P(\rho)d\rho = 1, \tag{3}$$

where the measure $\rho$ can be $R_\tau$, $r_\tau$, phase, or phase difference defined in the later discussions.

The probability distributions of the normalized returns $r_\tau(t)$ with different time scale $\tau$ for the DJIA are shown in Fig. 3. Here we have separated intraday data from interday data in the analyses and also compare the interday data with annual periodicity.

In Fig. 3(a), probability distributions of the normalized returns for intraday data with different time sampling intervals $\tau = 10, 30, 130, 390$ min are shown. Figure 3(b) is a comparison of the probability distributions of normalized returns for time sampling intervals from 10 min to 1 week. Probability distributions of the normalized returns with a time sampling interval of 10 min over different periods of the DJIA are shown in Fig. 3(c). Note that fluctuation strengths of probability distributions for different time scales are associated with the number of data points, and here the longer the time scale is, the fewer is the number of data points. From these figures, it is clear that $P(r_\tau)$ is independent of time sampling intervals and periods. In other words, with a proper

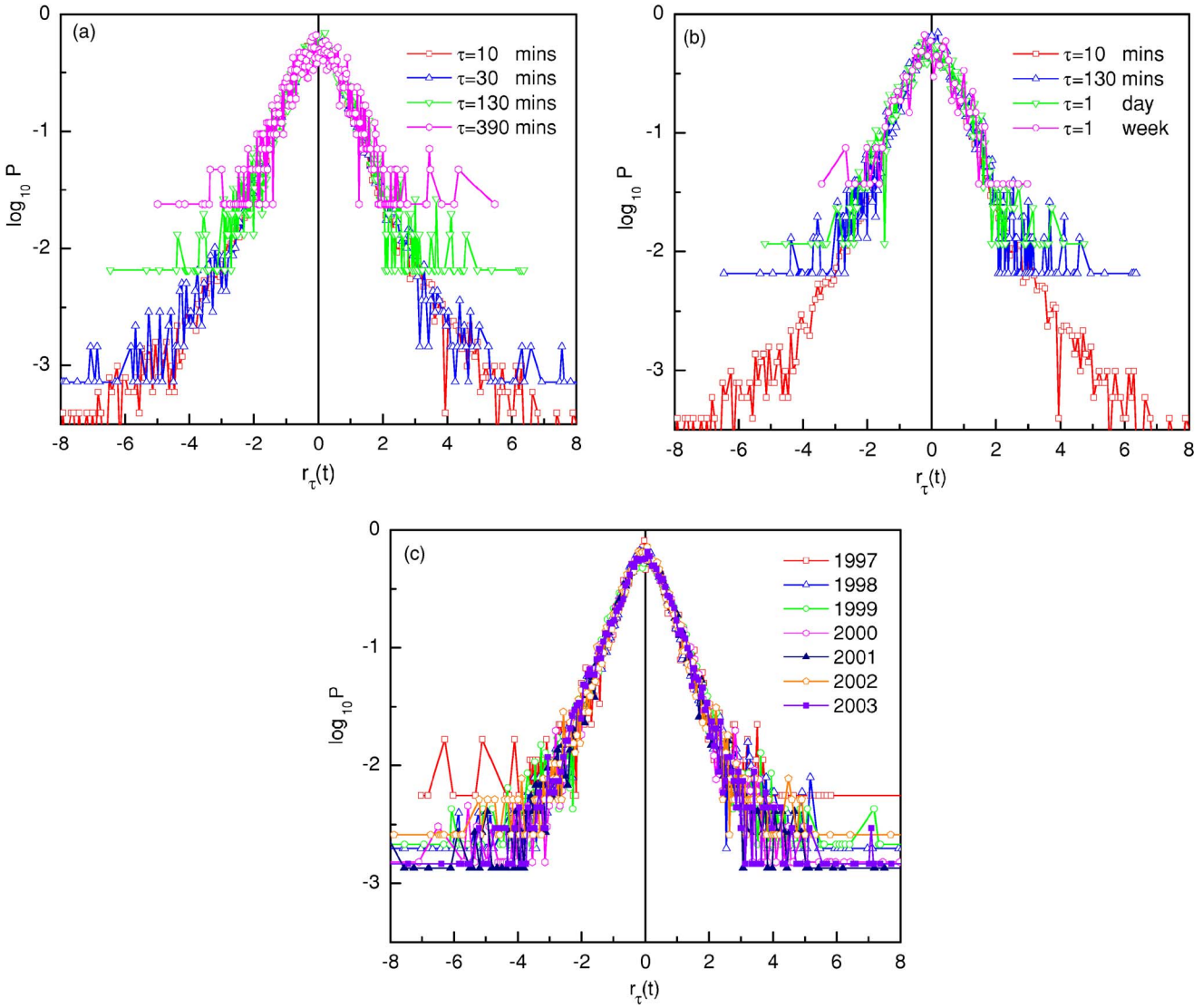


FIG. 3. (Color online) Probability distributions $P(r_\tau)$ of normalized returns of the DJIA, $r_\tau(t)$, (a) for intraday data with time sampling intervals of multiples of 10 min, (b) with a time sampling interval from 10 min to 1 week, and (c) with a time sampling interval of 10 min for different periods ranging from 1997 to 2003.

self-normalized factor, time series with different time scales can be view as a single one. The same phenomenon is applicable to $r_\tau(t)$ for NASDAQ returns, and we shall not repeat the discussion to save space. The existence of scaling behavior is then clear.

It was reported that the probability distributions of the normalized returns can be well described by the so-called double-exponential law (also known as the Laplace distribution) $P(r_\tau) \sim \exp(-|r_\tau|/\kappa)$, where $\kappa$ is a constant [22,23]. The double-exponential distribution of return at not-too-long times $t$ is a universal, ubiquitous feature of financial time series and was observed for different countries, stock-market indices, and individual stocks [23]. According to Ref. [23], the central part of the curves shown in Fig. 3 can be fitted by the scaling form using a Bessel function, where 99% of probability reside and statistics is good, followed by power laws in the far tails, where data statistics is often poor. These features are well caught by the Heston stochastic process [24]. For detailed discussions of the exponential-to-Gaussian crossover see Ref. [23].

## IV. SCALING ANALYSIS

According to scaling ansatz, a scaling theory for a system can be established provided that there is an extensive quantity in the system. In Fig. 3, time series with various sample time scales can be rescaled by a self-normalized factor implicitly depending on the corresponding time scales. Therefore, for financial time series, the extensive quantity is the time sampling interval $\tau$. An important issue we should notice here is the problem of data treatment in mixing intraday data and interday data. Since there are ambiguities in this issue, we perform scaling analysis both on intraday data [case (A)] and mixture of intraday and interday data [case (B)].

For case (A), data with certain time scale $\tau$ are sampled with fixed time sampling intervals of trading time from in-

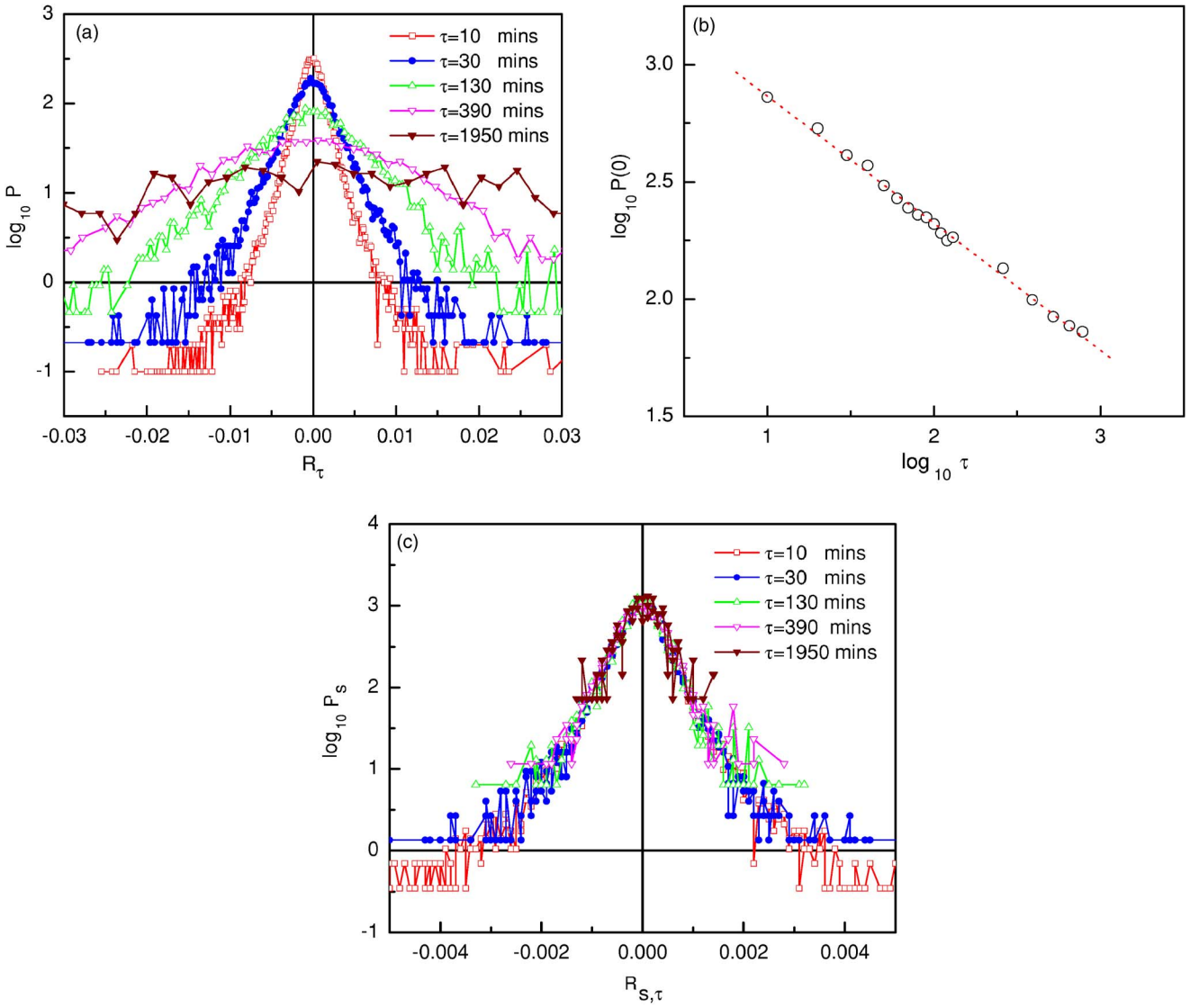


FIG. 4. (Color online) (a) Probability distributions $P(R_\tau)$ of the return variations of the DJIA, $R_\tau(t)$, for interday data observed at time intervals $\tau$. (b) Probability of return variation $P(R_\tau(t)=0)$ as a function of the time sampling intervals $\tau$. The slope of the best-fit straight line is $-0.54\pm0.01$. (c) Scaled plot of the probability distributions shown in (a) with $\alpha=1.84$.

traday data. In the case of a trading time is less than 1950 min in a week, the time interval of 1950 min may spread more than 1 week. In this analysis, there is no ambiguity in time sampling intervals, but may has uncertainty in mixtures of interday data.

For case (B), time sampling intervals are 10, 30, and 130 min for intraday data, 1 day for daily data, and 1 week for weekly data. In this analysis, the time sampling intervals of trading time is not fixed.

We first examine case (A). Figure 4(a) shows the probability distributions $P(R_\tau)$ of the intraday frequency variations of the DJIA with $R_\tau(t)$ observed at five different time intervals $\tau$, ranging from 10 to 1950 min. In contrast with those for normalized returns shown in Fig. 3, the probability density functions for returns $R_\tau(t)$ with different time sampling intervals do not converge to a single curve. However, according to Refs. [9,2], it is possible to make different$P(R_\tau)$ converge to a single curve by performing a scaling analysis. The simplest method is that we first *shift* these curves to make their maxima overlap and then rescale time intervals if necessary. To achieve this, we plot $P(R_\tau=0)$ with respect to the time sampling intervals $\tau$ in Fig. 4(b). The distributions of $P(R_\tau=0)$ with respect to $\tau$ plotted in logarithmic scale are linear. Accordingly, we take $P(R_\tau=0)$ with respect to $\tau$ from $\tau=10$ minutes to 780 min. The best fitting straight line is also plotted in Fig. 4(b), and it obeys

$$\log_{10}P(R_\tau=0) = C - \frac{1}{\alpha}\log_{10}\tau, \tag{4}$$

where $C$ is a constant. By measuring the slope of the fitting straight line, we have $\alpha=1.84\pm0.03$ which is larger than 1.4 in Ref. [9], but is consistent with $\alpha\leqslant 2$, the condition for stable Lévy distributions [3]. We then rescale returns $R_\tau(t)$ and the probability density function $P(R_\tau)$ according to [9]

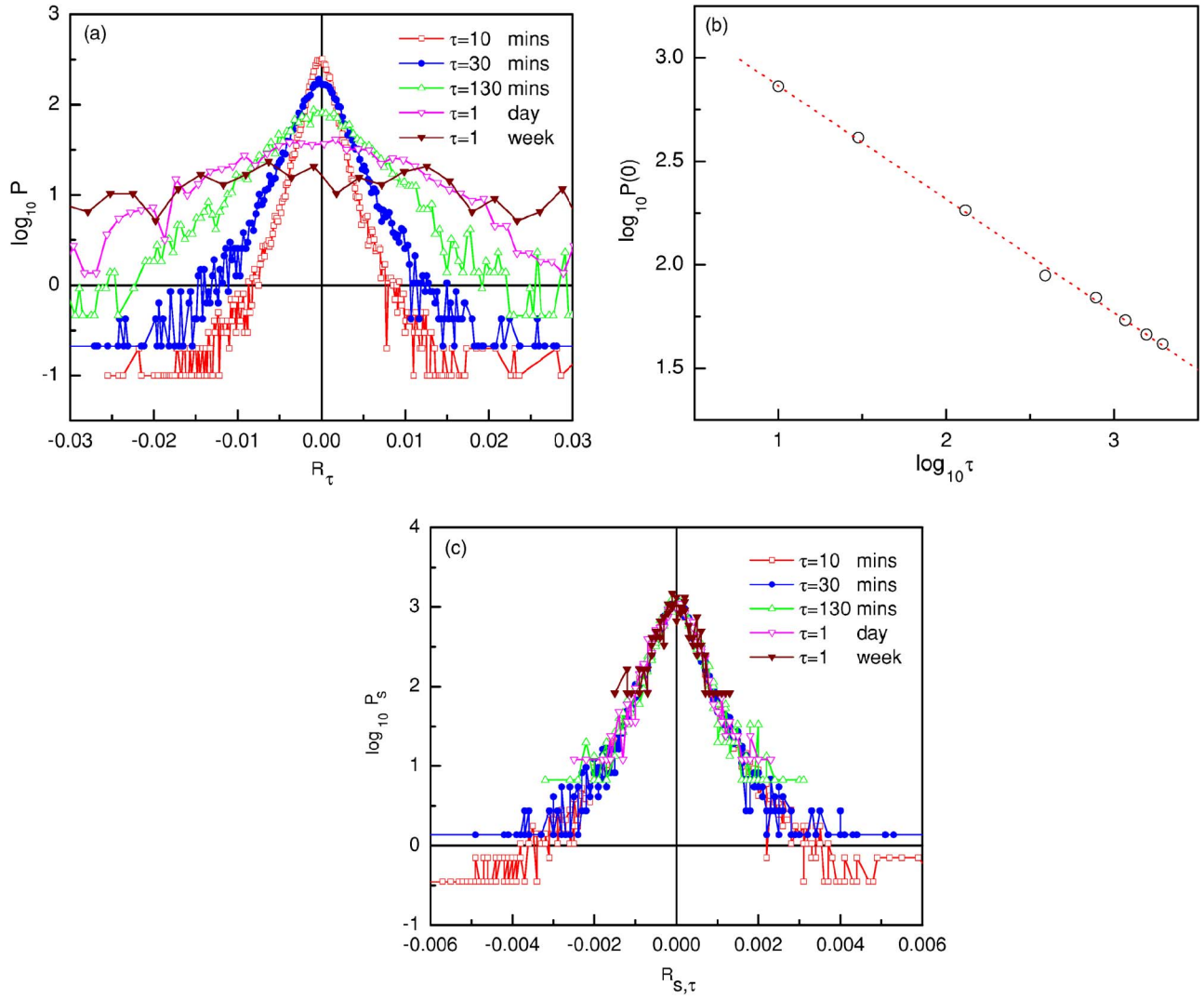


FIG. 5. (Color online) (a) Probability distributions $P(R_\tau)$ of the return variations of the DJIA, $R_\tau(t)$, for interday and interday data observed at time intervals $\tau$. (b) Probability of return variation $P(R_\tau(t)=0)$ as a function of the time sampling intervals $\tau$. The slope of the best-fit straight line is $-0.55\pm0.01$. (c) Scaled plot of the probability distributions shown in (a) with $\alpha=1.82$.

$$R_{s,\tau} = \frac{R_\tau}{\tau^{1/\alpha}}, \tag{5}$$

$$P_s(R_{s,\tau}) = \frac{P(R_\tau)}{\tau^{-1/\alpha}}. \tag{6}$$

Figure 4(c) shows a scaled plot of the probability distributions with $\alpha=1.84$. Probability distributions of time scales with properly scaled can coincide with each other very well.

Next, we examine case (B). We perform the same analysis on the intraday data and interday data (daily and weekly), and the results are shown in Fig. 5. The value of $\alpha$ is $\alpha=1.82\pm0.03$, which is about 1.2% smaller than that for intraday data. The small difference shall be due to mixture and nonmixture of intraday and interday data. A more rigorous examination of the difference can be investigated by the introduction of an effective overnight time lag such as the analysis done in Ref. [23]. According to our analyses herein, there is no significant difference between cases (A) and (B). We will then focus on the analysis of intraday data hereinafter.

For NASDAQ index data, we perform the same analysis of case (A) and the results are shown in Fig. 6. The value of $\alpha$ is $\alpha=1.75\pm0.03$, which is about 4.9% smaller than that of intraday data of DJIA returns and is also smaller than 2, which indicates returns of the NASDAQ index belong to stable Lévy distributions as well. Furthermore, the scaling behavior is also well described by $\alpha=1.75$ as shown in Fig. 6(c).

Here we note that if the index change $Y(t)-Y(t-\tau)$ is used for the scaling analysis, we get $\alpha=1.71\pm0.05$ [case (A)] and $1.77\pm0.08$ [case (B)] for the DJIA and $\alpha=1.51\pm0.04$ for NASDAQ. The significant difference of the value $\alpha$ for NASDAQ is due to the fact that during the considered period, changes of the NASDAQ index changed significantly such that the assumption of tick-by-tick data is linear is incorrect. This fact demonstrates the conclusion of Ref. [21] that returns should be used instead of index change for the scaling analysis. Furthermore, Eqs. (5) and (6) can be sum-

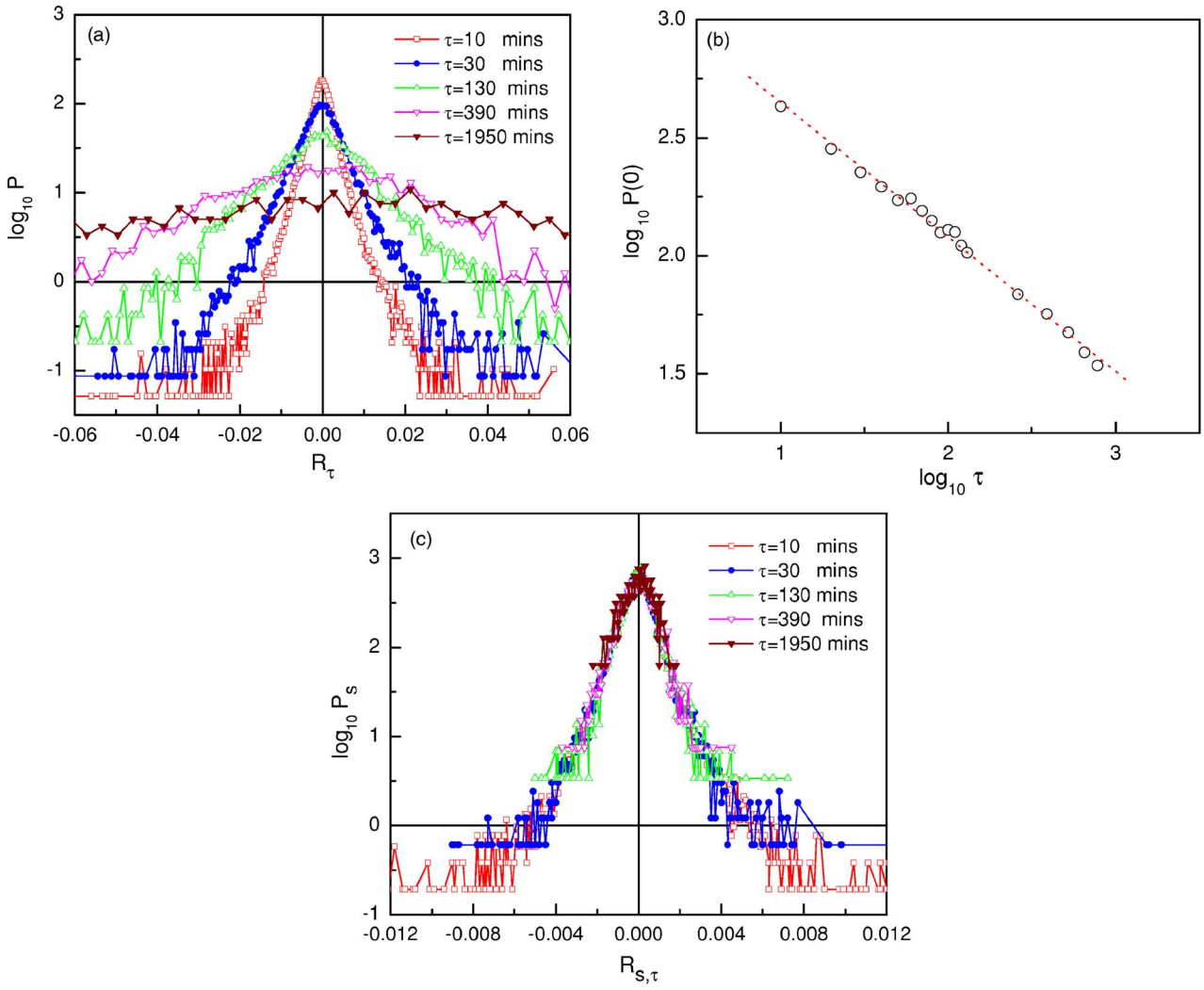


FIG. 6. (Color online) (a) Probability distributions $P(R_\tau)$ of the return variations of NASDAQ, $R_\tau(t)$, for interday data observed at time intervals $\tau$. (b) Probability of return variations $P(R_\tau(t)=0)$ of as a function of the time sampling intervals $\tau$. The slope of the best-fit straight line is $-0.57\pm0.01$. (c) Scaled plot of the probability distributions shown in (a) with $\alpha=1.75$.

marized as $P(R_\tau)\sim f(R_\tau/\tau^b)$, where $b=1/\alpha$. According to our analysis, $b=0.54$ and 0.55 for the DJIA and $b=0.57$ for NASDAQ, which are close to the value 0.5 found in Ref. [23], in which the function $f(R_\tau/\tau^b)$ is considered as the exponential function in short-time limit and the Gaussian function in long-time limit.

## V. PROBABILITY DISTRIBUTION OF THE INSTANTANEOUS PHASE

The analysis of scaling in Sec. IV is for a survey of statistical property of financial time series. In this section, we intend to investigate other nature of financial time series. However, as shown in Fig. 1, time series of index are in general nonlinear and nonstationary. The conventional method for nonstationary time series analysis based on Fourier transform, such as spectrum analysis, wavelet analysis, etc., may suffer from the limitation of linearity. This finding has been discussed elsewhere, and further review of the advantages, limits, and shortcomings of the existing time series analysis methods can be found in Ref. [16]. In view of these, here we suggest an approach which is based on the concept that an instantaneous phase can catch the characteristic features of financial time series. The idea originates from the fact that phases of a time series usually contain rich information about the structures of the time series. The proposal will be very useful if such information can be extracted faithfully in further analyses. Therefore, to achieve this, we introduce the Hilbert-Huang time signal analysis method [16], which is suitable for the analysis of nonstationary time series, to define and calculate the instantaneous phase.

The Hilbert-Huang method of time signal analysis consists of the so-called empirical mode decomposition and the Hilbert spectral analysis. The EMD method is developed from the assumption that any time series consists of simple intrinsic modes of oscillation, and the essence of the method is to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically and then decompose the data accordingly [16]. This is achieved by sifting

data to generate IMF's. The IMF's introduced by EMD are a set of well-behaved intrinsic modes, and these functions satisfy the conditions that they are symmetric with respect to the local zero mean and have the same numbers of zero crossings and extremes. Therefore, the Hilbert transform can be directly used to calculate the instantaneous phase after the decomposition processes.

The algorithm to create IMF's in EMD is rather elegant, and it mainly consists of two steps. First, the local extremes in the return time series data, $R_\tau(t)$, are identified. Then, all the local maxima are connected by a cubic spline line $U(t)$, which forms the upper envelope of the time series. At the same time, the same procedure is applied for the local minima to produce the lower envelope, $L(t)$. Both envelopes will cover all the original time series. The mean of upper envelope and lower envelope, $m_1(t)$, given by

$$m_1(t) = \frac{U(t) + L(t)}{2}, \tag{7}$$

is a running mean. We then subtract the running mean $m_1(t)$, from the original time series $R_\tau(t)$, and get the first component $h_1(t)$,

$$R_\tau(t) - m_1(t) = h_1(t). \tag{8}$$

The resulting component $h_1(t)$ is an IMF if it satisfies the following conditions: (i) $h_1(t)$ is free of riding waves. (ii) It displays symmetry of the upper and lower envelopes with respect to zero. (iii) The numbers of zero crossing and extremes are the same or only differ by 1. If $h_1(t)$ is not an IMF, the sifting process has to be repeated as many times as is required to reduce the extracted signal to an IMF. In the subsequent steps of sifting process, $h_1(t)$ is treated as the data,

$$h_1(t) - m_{11}(t) = h_{11}(t). \tag{9}$$

Again, if the function $h_{11}(t)$ does not yet satisfy criteria (i)–(iii), the first sifting process continues up to $k$ times until some acceptable tolerance is reached and

$$h_{1(k-1)}(t) - m_{1k}(t) = h_{1k}(t). \tag{10}$$

If the resulting time series is the first IMF, then it is designated as $c_1 = h_{1k}(t)$. The first IMF component from the data contains the highest oscillatory frequency found in the original data $R_\tau(t)$.

Subsequently, the first IMF is subtracted from the original data and the difference $r_1$, given by

$$R_\tau(t) - c_1(t) = r_1(t), \tag{11}$$

is a residue. The residue $r_1(t)$ is taken as if it were the original data, and we apply to it again the sifting process. Following the above procedures, the process of finding more intrinsic modes $c_i$ continues until the last mode is found. The final residue will be a constant or a monotonic function which represents the general trend of the time series data. Finally, we get

$$R_\tau(t) = \sum_{i=1}^{n} c_i(t) + r_n(t), \tag{12}$$

$$r_{i-1}(t) - c_i(t) = r_i(t), \tag{13}$$

where $r_n$ is a residue. In general, $r_n$ is a constant or a monotonic function which represents the general trend of the time series.

To perform the EMD method on a financial time series, one may or may not impose an intermittency as an additional condition in the sifting process, depending on the nature of the financial time series under consideration. The intermittency can be considered as a window used to eliminate the end effects and to facilitate computation. However, a characteristic intermittency in trading time of a stock market may be indefinite. In particular, for a truly nonstationary process like index (return) time series, there is no time scale to guide the choice of window size. Therefore, here we do not impose definite intermittencies in the sifting process. Hence, in the sifting process, the structures of the time series with primary time sampling intervals are closely preserved in the first mode.

We take intraday returns $R_\tau(t)$ with time sampling interval of 10 min as the primary time series and then perform EMD to decompose $R_\tau(t)$ into 14 IMF's. The results are shown in Fig. 7(a), in which only the first three IMF's are shown. The physical meanings of the decomposition are clear from the features of IMF's. Let us first compare time series $R_\tau(t)$ and IMF's $c_1$ and $c_2$ in Fig. 7(a). According to Eqs. (12) and (13), $R_\tau(t)$ consists of 14 IMF's and each IMF is independent from the others. The term "independent" here is in some sense equivalent to the term "orthogonal" in the theory of finite-dimensional vector space. In other words, each IMF cannot be represented by other IMF's decomposed from the same primary time series. The main difference between IMF's $c_1$ and $c_2$ is the intermittencies they own. IMF $c_1$ is the first mode separated from $R_\tau(t)$ after the sifting process, and it has the highest frequency among 14 IMF's. Since no criterion is imposed on the intermittency, there is no specified relation between the intermittencies of $c_1$ and $c_2$. Furthermore, if one IMF dominantly catches characteristic features of $R_\tau(t)$, then its contribution is distinguishable in an observation like Fig. 7(a). It is obvious that $c_1$ catches the main structures of $R_\tau(t)$ since the time series of $R_\tau(t)$ is mainly characterized by its highest-frequency component. However, we should note that this is case by case and the conclusion may not be applicable to other time series.

In our analysis, it is very important to note that IMF $c_1$ is not equal to time series $R_\tau(t)$. If we evaluate some quantities specifically defined for $R_\tau(t)$ from $c_r$'s, the results may be quite different. Actually, it is not reasonable to copy all the fundamental statistics primarily performed on return $R_\tau(t)$ to IMF's.

After IMF's have been obtained from the EMD method, one can further calculate instantaneous phases of IMF's by applying the Hilbert transform to each IMF component—say, the $r$th component. The procedures of the Hilbert transform consist of calculation of the conjugate pair of $c_r(t)$—i.e.,

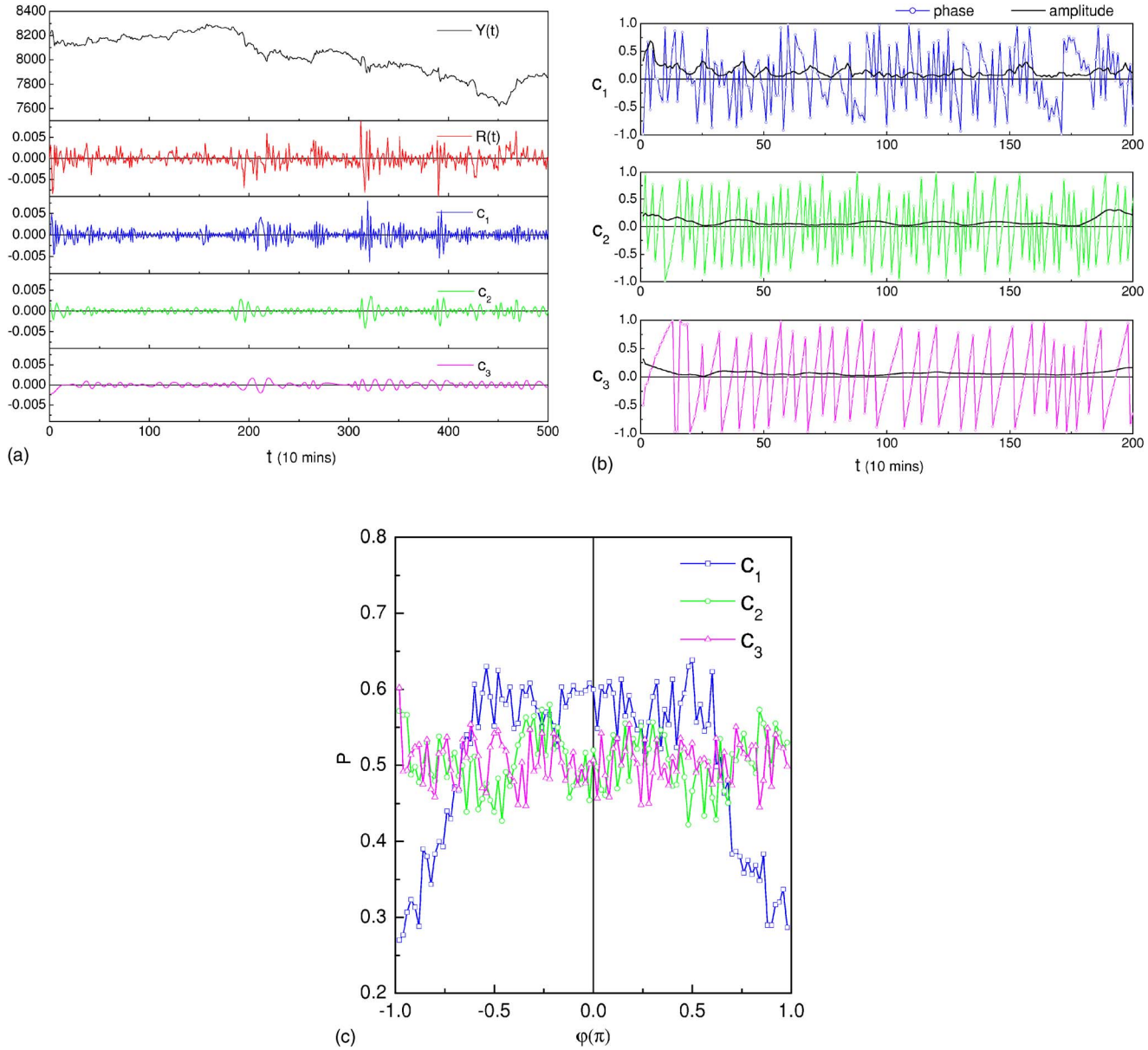


FIG. 7. (Color online) (a) Intraday DJIA index and the corresponding return sampled by10 min and the first three IMF's; (b) amplitude and phase variations of the IMF's in (a), in which the amplitude have been adjusted by multiplying 100; and (c) probability distribution of phases.

$$y_r(t) = \frac{1}{\pi} \mathrm{P} \int_{-\infty}^{\infty} \frac{c_r(t')}{t - t'} dt', \tag{14}$$

where "P" indicates the Cauchy principal value. With this definition, the two functions $c_r(t)$ and $y_r(t)$ forming a complex conjugate pair define an analytic signal $z_r(t)$:

$$z_r(t) = c_r(t) + i y_r(t), \tag{15}$$

which can also be expressed as

$$z_r(t) = A_r(t) e^{i\phi_r(t)}, \tag{16}$$

with amplitude $A_r(t)$ and the phase $\phi_r(t)$ defined by

$$A_r(t) = [c_r^2(t) + y_r^2(t)]^{1/2}, \tag{17}$$

$$\phi_r(t) = \arctan\left(\frac{y_r(t)}{c_r(t)}\right). \tag{18}$$

Then, we can calculate the instantaneous phase according to Eqs. (14) and (18). The amplitudes and phases of the IMF's in Fig. 7(a) calculated by the Hilbert transform are shown in Fig. 7(b). Note that the magnitude of the amplitude is significantly smaller than the phase. To show both amplitude and phase in the same figure, here we have adjusted the magnitude of amplitude by multiplying by 100 to make it fit to the same scale for phase.

Similarly, we can perform EMD on the time series with time sampling interval of 30, 130, and 390 min and the phase distributions are shown in Figs. 8(a)–8(c), respectively. For

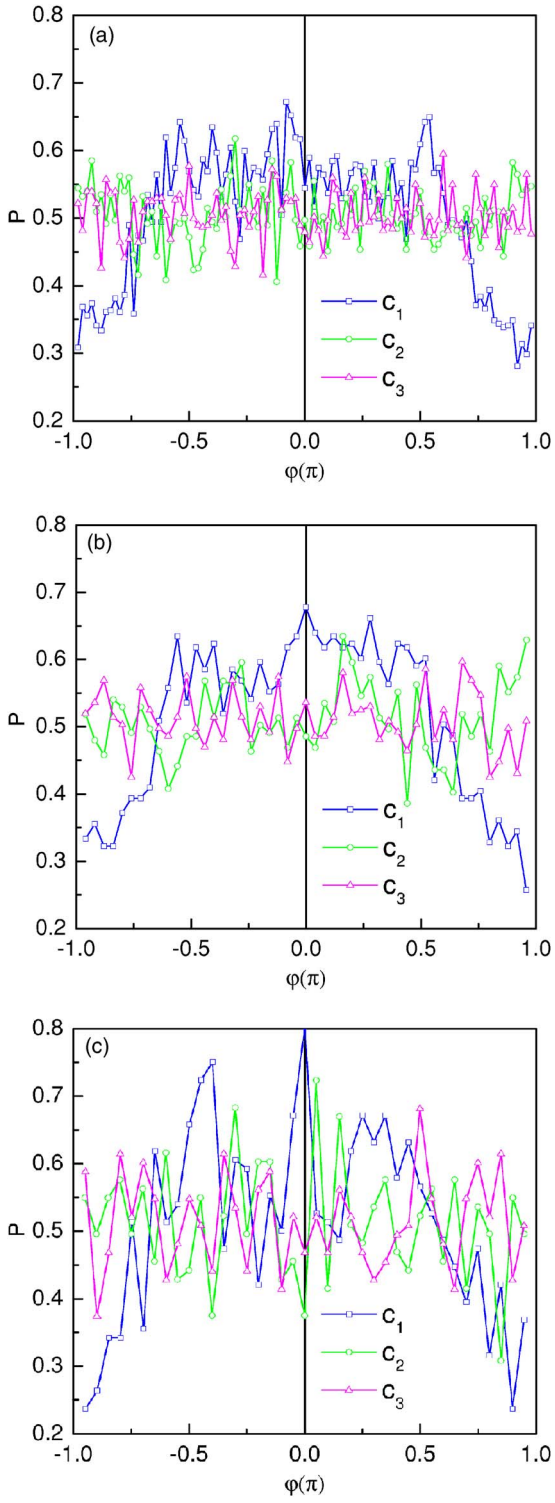


FIG. 8. (Color online) Probability distributions of phases of the first three IMF's for intraday returns of the DJIA index sampled by (a) 30 min, (b) 130 min, and (c) 390 min.

the time series with time sampling interval of 1 day and 1 week, similar patterns can be obtained. We find that except for the first IMF's of these time series, the phases of other IMF's are randomly distributed and have equal probabilities for all possible phases—i.e., $-\pi \leqslant \phi \leqslant \pi$. Note that for time series with too low frequency such as weekly, the number of sampling data points are too few to exhibit this feature explicitly. Figures 9(a) and 9(b) show the amplitude and phase distributions of the first IMF's of these time series, respectively. The probability density functions of amplitudes for the first IMF's are general Boltzmann distributions. Among these, the phase distribution is quite interesting. Most phases of the IMF's locate at $-0.5\pi \leqslant \phi \leqslant 0.5\pi$. For clarity, we locate the index with instantaneous phase $-0.5\pi \leqslant \phi \leqslant 0.5\pi$ on the time series of the DJIA index with red spots, and the results are shown in Fig. 9(c). The patterns of red spots do not have particular rules and are nonuniform distributed. The clustered distribution of phase originates from the abruptly changing behaviors of the index time series, which is a nature of time series with intermittency close to the sample time scale $\tau$. We find that these behaviors exist in all sample time scale (time sampling intervals of multiples of 10 min) of intraday data and are believed to persist in interday time scales (daily, weekly, and even lower frequencies). From another point of view, the behaviors of abruptly change imply nonpredictable and stochastic features of the index. These features may be understood by the stochastic volatility model which is a log-Brownian model with random diffusion coefficients [25,26]. In particular, it has been reported that essential features of stock price dynamics can be well modeled by a number of stochastic volatility models [27].

We further preform analysis on the NASDAQ index time series under the same framework, and the results are shown in Fig. 10. Both the probability distributions of the first IMF's of returns of DJIA and NASDAQ indices are Boltzmann distributions, except for a difference in scale. It is remarkable that the distributions of phases are the same, which implies that it is a characteristic behavior of this kind of time series. As mentioned above, the behavior indicates nonpredictable features of index time series and is very different from regular signals or pseudoregular signals. For instance, a typical time series of respiratory cycles [28] is shown in Fig. 11(a). In general, the respiratory cycle is not a regular time series but is disturbed by body actions and noises. The third IMF $c_3$ obtained by EMD catches the main structures of this time series [29], and the corresponding probability distribution of amplitudes is shown in Fig. 11(b). We compare the probability distributions of phases for returns of DJIA and NASDAQ indices, foreign exchange [30], and respiratory time series in Fig. 11(c). From this figure, that return time series and respiratory time series belong to different classes is quite apparent. The same analysis can also be performed on other time series, such as temperature variation, population, etc. The investigations will be reported elsewhere [31].

The different patterns of phase distributions for the return time series and the respiratory time series can be understood from the mechanisms and the sampling rate of the processes. It is recognized that financial markets belong to the self-organized system [32], which shows a nonequilibrium steady state of an extended system with a steady drive, but irregular burst like relaxations [33]. The financial markets thus can be modeled by a random process with stochastic volatility. There is no sinusoidlike cyclic rhythm existing in the time scale catching characteristic structures of high-frequency time series and can be used to define as a characteristic time scale. The definition of returns then makes the wave form of

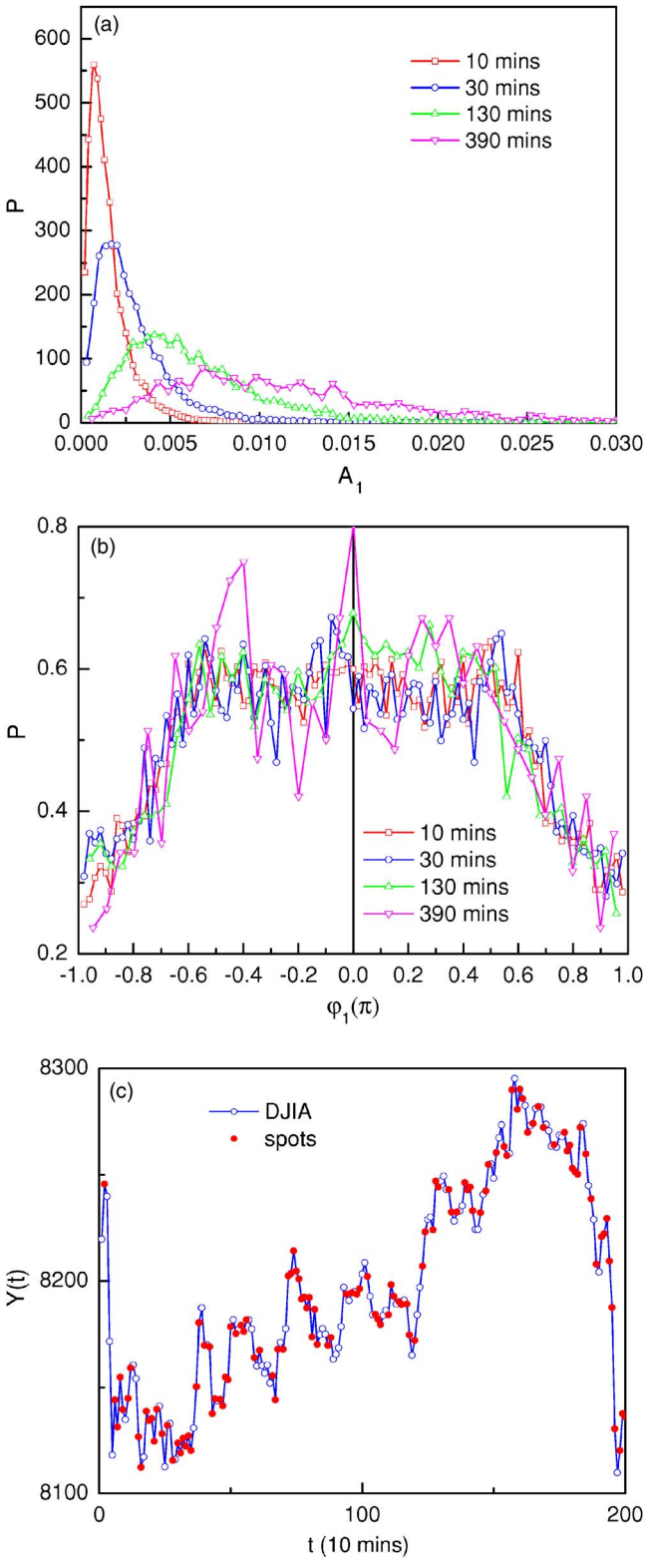


FIG. 9. (Color online) Probability distributions of (a) amplitudes and (b) phases for the first IMF's of the returns of the DJIA index sampled by 10, 30, 130, 390 min; (c) intraday DJIA index sampled by 10 min and with red spots indicating the instantaneous phase $-0.5\pi \leqslant \phi \leqslant 0.5\pi$.

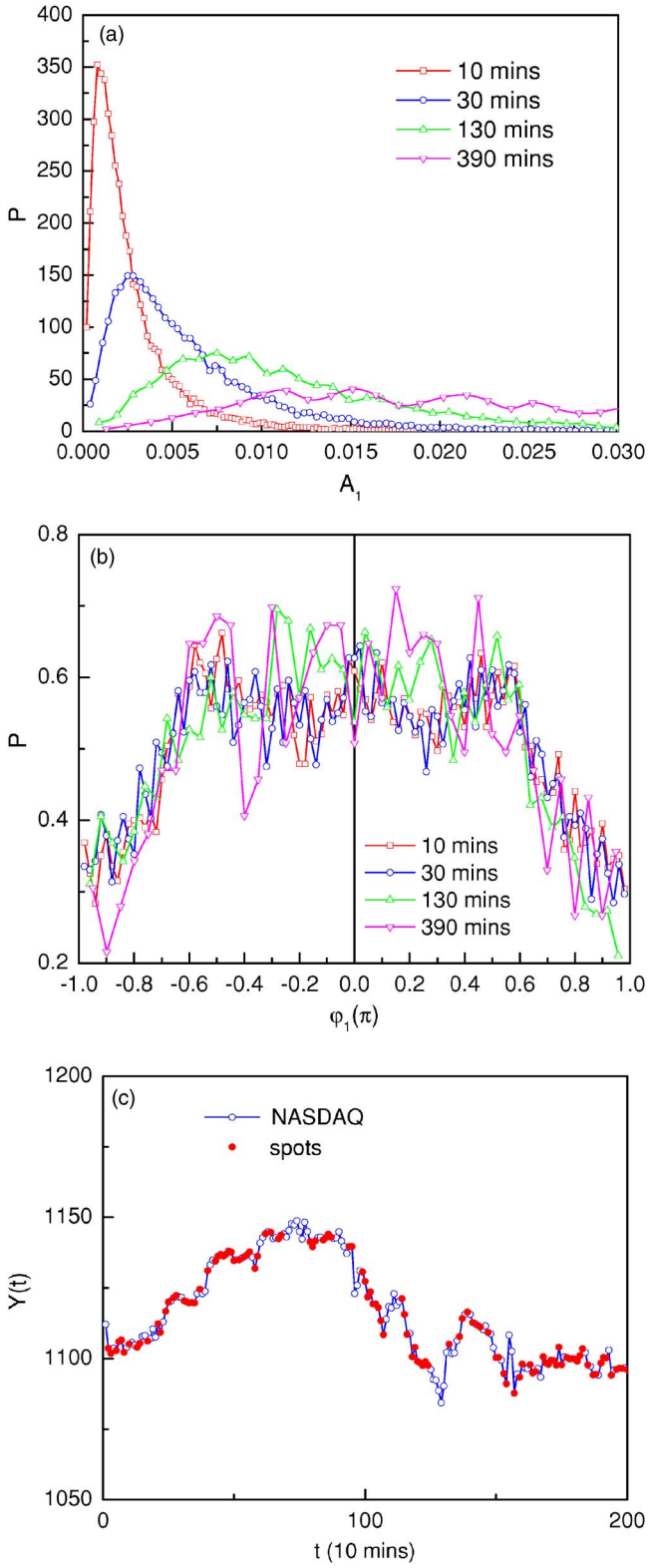


FIG. 10. (Color online) Probability distributions of (a) amplitudes and (b) phases for the first IMF's of the returns of the NASDAQ index sampled by 10, 30, 130, 390 min; (c) intraday NASDAQ index sampled by 10 min and with red spots indicating the instantaneous phase $-0.5\pi \leqslant \phi \leqslant 0.5\pi$.

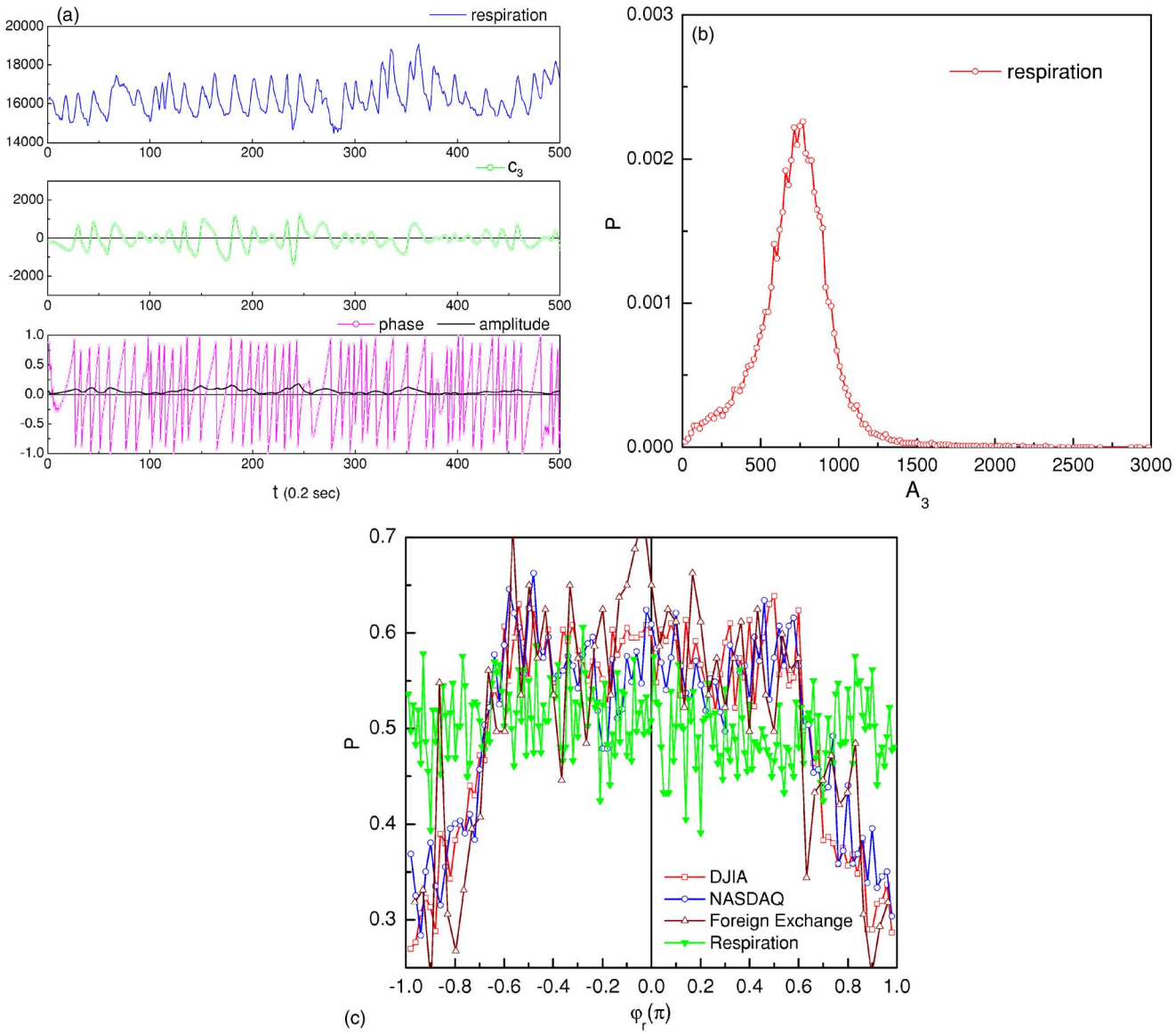


FIG. 11. (Color online) (a) Typical human respiratory time series [28] and the third IMF, which catches the main structures and the corresponding amplitude and phase variations; (b) probability distribution of amplitudes; and (c) probability distributions of phases for the first IMF's of the returns for the DJIA and NASDAQ indices sampled by 10 min, daily returns of foreign exchange [30], and the third IMF of the respiratory time series in (a).

the corresponding time series in a zigzag fashion. As a result, the intermittency of return time series always has the order of the sampling time intervals. In contrast to the financial time series, the respiratory time series is a measurement of respiratory cycles which are controlled by neural and physiological systems, and are also influenced by mechanical effects. The respiratory signals may represent measures of the volume of expansion of ribcage and can be described by sinusoidal wave forms with time-varying amplitude and frequency [29]. Since respiration always completes a cycle in a definite time period, the characteristic time scale can be defined in this system. For example, each respiratory cycle takes 2–6 sec depending on the physiological situations and age, and a characteristic time scale can then be defined accordingly. Therefore, the time series of respiratory cycle can be sampled by sufficient short-time sampling intervals such that the structures of the wave form can be precisely caught. It follows that the calculation of phase distribution for the wave form leads to a homogeneous pattern in a range from $-\pi$ to $\pi$.

## VI. CORRELATION BETWEEN THE DJIA AND NASDAQ INDICES

At first glance at Fig. 1, the DJIA and NASDAQ indices seem to have explicit correlations in several epoches. For example, two indices decline in August and September of 1998 and in March of 2001. Two indices abruptly decline in September of 2001 due to the accident of 9/11 attack. These big changes are in-phase, and there are also out-of-phase changes, such as those in the period from February to March

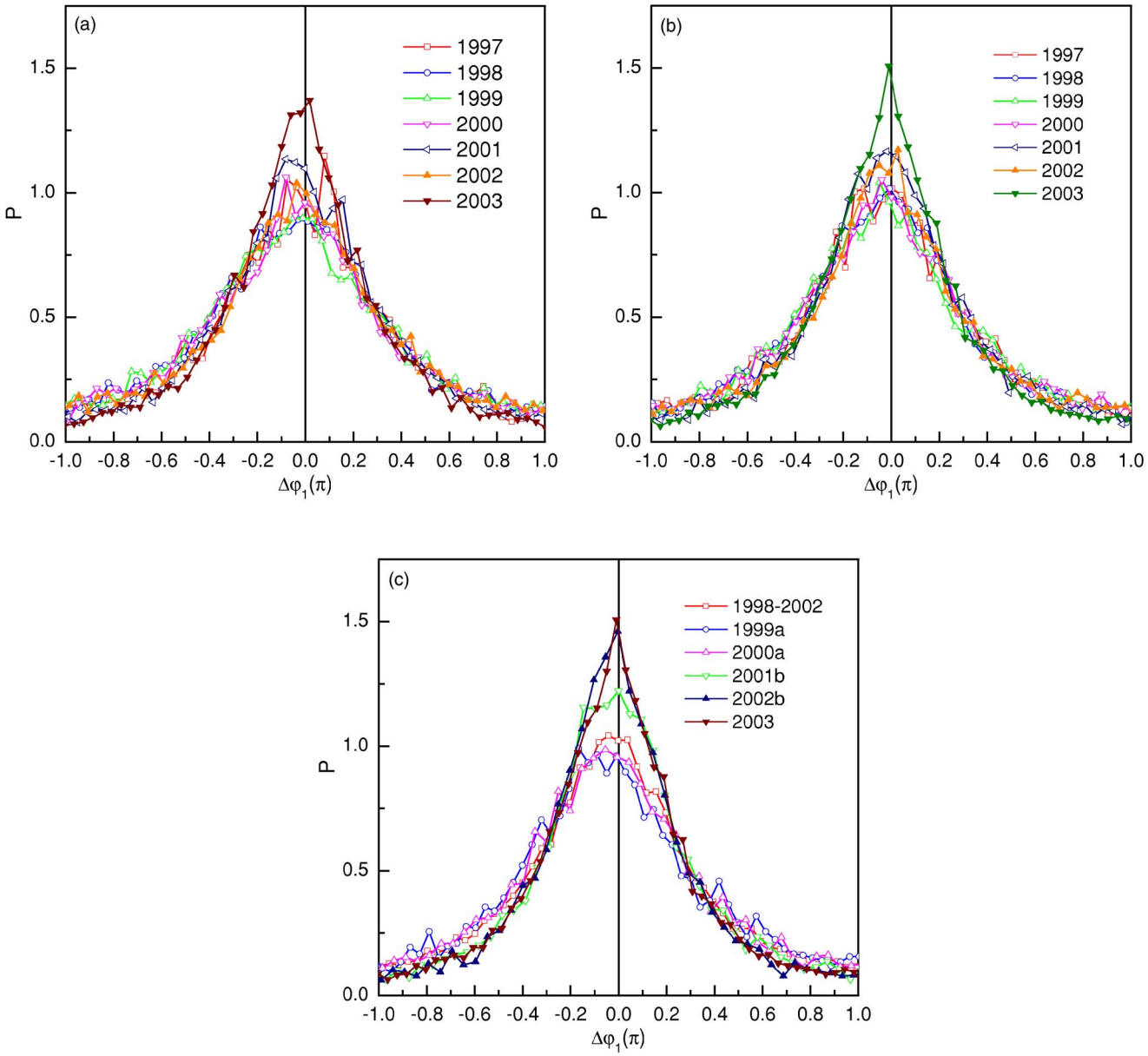


FIG. 12. (Color online) Probability distributions of phase differences between the first IMF's of (a) DJIA and NASDAQ indices and (b) returns of DJIA and NASDAQ indices for different periods ranging from 1997 to 2003, and (c) returns of DJIA and NASDAQ indices for the years 1998–2002 and for certain periods and events.

of 2000. In other periods, we can also find similar behaviors in shorter time scales.

To investigate the correlative behaviors set forth, here we also apply the Hilbert-Huang method to calculate the instantaneous phases of several epoches of the index and return time series of the DJIA and NASDAQ indices. To be statistical meaningful, each epoch will have more than 3000 sampling points. Here we further define phase differences of the first IMF's for different indices. Take the DJIA as a reference and define the phase difference $\Delta\phi_r$ as

$$\Delta\phi_r = \phi_r(\text{NASDAQ}) - \phi_r(\text{DJIA}), \tag{19}$$

and then calculate the probability distributions for various periods, in units of years. The results are shown in Figs. 12(a) and 12(b) for index time series and return time series, respectively. The skewness and kurtosis of the corresponding statistics are summarized in Table I. It is interestingly that, in the year 2003, both Figs. 12(a) and 12(b) have sharp peaks around zero of the phase difference compared with other periods, indicating that the relation of phases between two indices is closer to each other in the year 2003. This implies more correlative behaviors between two indices in the year 2003. Suppose this is a general trend; the stronger correlative behaviors may in some sense provide implications to market investors in buying and selling trading strategy [17].

Furthermore, in spite of the small differences between the statistics based on the index [Fig. 12(a)] and return [Fig. 12(b)] in Table I, a general feature shall be the negative value of skewness for the years 2001 and 2002 which indicates that

TABLE I. Skewness and kurtosis of the phases differences between the first IMF's of (a) DJIA and NASDAQ indices, and (b), (c) returns of DJIA and NASDAQ indices, for certain periods and events. The corresponding distributions are, respectively, shown in Figs. 12(a)–12(c).

| | | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|---|---|---|
| (a) | Skewness | 0.006 | 0.072 | 0.101 | 0.067 | −0.028 | −0.034 | 0.042 |
| | Kurtosis | 0.888 | 0.653 | 0.567 | 0.669 | 1.274 | 0.738 | 1.834 |
| | | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
| (b) | Skewness | −0.004 | 0.079 | 0.069 | 0.115 | −0.025 | 0.003 | 0.026 |
| | Kurtosis | 0.907 | 0.843 | 0.728 | 0.779 | 1.326 | 0.934 | 1.990 |
| | | 1998-2002 | | 1999a | 2000a | 2001b | 2002b | 2003 |
| (c) | Skewness | 0.050 | | 0.097 | 0.133 | −0.050 | −0.001 | 0.026 |
| | Kurtosis | 0.903 | | 0.776 | 0.848 | 1.556 | 2.104 | 1.990 |

the peaks of distributions are deviate from zero and are slightly shifted to a negative domain. This implies that the phase of the DJIA is ahead of NASDAQ on average for the years 2001 and 2002. In other words, trading activities in the DJIA have an affect on NASDAQ more than NASDAQ has an affect on the DJIA. However, in other periods, trading activities of NASDAQ have an affect on the DJIA more than the DJIA has an affect on NASDAQ. This seems to be natural, since the DJIA is more stable and mature than NASDAQ in composition and the stocks in NASDAQ are usually more active than those in the DJIA. Consequently, in the years 2001 and 2002 in which stock markets were influenced by the event of the 9/11 attack, the performance of mutual stocks of the DJIA greatly affected those of NASDAQ. In other periods such as the years 1998, 1999, and 2003, active stocks in NASDAQ then affected the environment of the stocks in the DJIA.

We further investigate distributions of phase differences for certain epoches and events. Figure 12(c) shows the probability distributions of phase differences between the first IMF's of returns of two indices over the years 1998–2002 and the first half year of 1999 (indicated by 1999a), the first half year of 2000 (indicated by 2000a), the last half years of 2001 and 2002 (indicated by 2001b and 2002b, respectively), and the whole year of 2003. We find that there is a remarkable change in the behavior of trading activities both in the DJIA and NASDAQ since the event of the 9/11 attack. More specifically, Fig. 12(c) and kurtosis statistics in Table I show that there were more correlative activities after 9/11 such that the distribution functions of 2001b and 2002b were quite different from those before 9/11. Note that there was a similar spectrum in the year 2003, which implies that the scenario persisted in later trading activities. This may be interpreted by faster communications and stronger event dependence after 9/11 in stock markets. In other words, the behaviors of the two indices became more correlative due to influences from common factors, such as news reports and events, systemic risk, macroeconomic announcement, and federal government policy. As a result, investors had similar trading strategies during the anniversary of 9/11 in the year 2002 under these influences.

## VII. CONCLUSIONS

In conclusion, we have investigated the scaling analysis, phase distribution, and phase correlation of DJIA and NASDAQ indices based on high-frequency intraday data.

A scaling analysis was performed both on DJIA and NASDAQ returns. For the DJIA index, the values of $\alpha$ are 1.84 and 1.82 for intraday and a mixture of intraday and interday cases, respectively. The small difference is due to a mixture and nonmixture of intraday and interday data within weekly frequency. Therefore, there is no remarkable difference between intraday and interday time sampling schemes for those with sampling frequency higher than weekly. For NASDAQ returns, the value of $\alpha$ for intraday data is 1.75, which is slightly smaller than that for DJIA returns. However, both DJIA and NASDAQ returns satisfy the stable Lévy distributions with $\alpha \leqslant 2$ [3]. The analysis of scaling also show that the two returns have nice scaling behaviors with respect to various time scales within a truncated time scale, which is consistent with the existing literature [9,21,23].

We further employed the Hilbert-Huang method of time signal analysis to define the instantaneous phase to catch characteristic features of index and return time series. The EMD method was used to decompose return time series into several IMF's, and the Hilbert transform was used to calculate the instantaneous phase of the first three IMF's accordingly. We find that except for the first IMF's of these time series which have phases mainly distributed within a range of $-0.5\pi \leqslant \phi \leqslant 0.5\pi$, the phases of other IMF's are randomly distributed and have equal probabilities for all possible phases. This behavior exists in all sample time scales (time sampling intervals of multiples of 10 min) of intraday data and interday data less than the weekly frequency. We expect that the same behavior also exists on a larger time scale. The phase distributions corresponding to abruptly changing behaviors indicate nonpredictable and stochastic features of the indexes. Furthermore, our results show explicitly that the phase spectrum of return time series falls into a class different from other signals, such as a time series of human respiration.

The investigations of correlations between DJIA and NASDAQ indices by the phase difference for various periods

and epoches show a remarkable picture of trading activities. We find that the phases of two index and return time series became closer in the year 2003 than in earlier years, and the phase of return time series for the DJIA index was ahead of that for the NASDAQ index in the years 2001 and 2002 on average. In other words, trading activities in the DJIA influenced NASDAQ more than NASDAQ influenced the DJIA in this period. This phenomenon was explained by the fact that the DJIA is more stable and mature than NASDAQ in composition, and the stocks in NASDAQ are usually more active than those in the DJIA. Consequently, in the years 2001 and 2002 in which stock markets are influenced by the event of the 9/11 attack, the performance of mutual stocks of the DJIA affected much those of NASDAQ. In other periods such as the years 1998, 1999, and 2003, active stocks in NASDAQ then affected the environment of the stocks in the DJIA.

Further, the phase distribution between two indices became closer after the event of 9/11. This implies an explicit change in the behavior of trading activities of the DJIA and NASDAQ after September 2001. A similar spectrum in the last half year of 2002 and the whole year of 2003 [Fig. 12(c)] further implies that the scenario persisted in later trading activities. This was interpreted by faster information transmission and stronger event dependence in stock markets after 9/11. In other words, two indices became more correlative due to influences from common factors, such as news reports and events, systemic risk, macroeconomic announcement, and federal government policy. Accordingly, investors had similar trading strategies during the anniversary of 9/11 in the year 2002 under the influence of anticipation.

It shall be heuristic to compare our method and the random matrix method [4–6]. The random matrix method is based on the concept of cross correlation of stocks, and the spectrum from market data [4,5] possesses a bulk of continuously distributed eigenvalues, which is similar to random matrix theory [4,34,35]. The method introduces a cross-correlation matrix to measure the statistical overlap of the fluctuations in the returns between pairs of stocks [4,5,36,37] and then solves the eigenvalues of the random matrix. The effects of the correlations [4,5,36,37] are manifest in the eigenvectors of those eigenvalues, and the corresponding patterns are related to the cooperative behavior in the fluctuations of the stock prices [38–40]. Therefore, the random matrix method is designated and suitable for the study of correlative behaviors in a collective system. It can only be applied to analyses with more than two stocks. In contrast to the random matrix method, in our method there is no assumption of correlations in the primary time series and the empirical decomposition of primary time series into IMF's is intuitive and straightforward. As shown in the analysis of this paper, our method can be used for the investigation of correlation and is also useful for the study of the intrinsic properties of an individual time series.

Finally, according to the impressive implications disclosed by our studies based on the concepts of phase distribution and phase correlation, we expect that our approach is also useful for statistical analysis of other time series, such as time series of physiological systems [29] and other social models [31]. Furthermore, it would also be interesting to test if our approach can be applied to the study of memory effects in financial time series [41].

## ACKNOWLEDGMENTS

This work was supported in part by Academia Sinica (Taiwan) under Grant No. AS-91-TP-A02 (M.-C.W.) and the National Science Council of the Republic of China (Taiwan) under Grant No. NSC 93-2112-M-033-005 (M.-C.H.). M.-C.W. acknowledges helpful discussions with Dr. Wen-Jong Ma.

[1] B. Zhou, J. Bus. Econ. Stat. **14**, 45 (1996).
[2] R. N. Mantegna and H. E. Stanley, *An Introduction to Econophysics, Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, England, 2000).
[3] J. Voit, *The Statistical Mechanics of Financial Markets*, 2nd ed. (Springer-Verlag, New York, 2003).
[4] L. Laloux, P. Cizeau, J. -P. Bouchaud, and M. Potters, Phys. Rev. Lett. **83**, 1467 (1999).
[5] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. Nunes Amaral, and H. E. Stanley, Phys. Rev. Lett. **83**, 1471 (1999).
[6] W. -J. Ma, C. -K. Hu, and R. E. Amritkar, Phys. Rev. E **70**, 026101 (2004).
[7] H. -C. Yu and M. -C. Huang, Appl. Financ. Econ. **14**, 1087 (2004).
[8] V. Plerou, P. Gopikrishnan, and H. E. Stanley, Nature (London) **421**, 130 (2003).
[9] R. N. Mantegna and H. E. Stanley, Nature (London) **376**, 46 (1995).
[10] K. Ohashi, L. A. Nunes Amaral, B. H. Natelson, and Y. Yamamoto, Phys. Rev. E **68**, 065204(R) (2003).
[11] J. F. Muzy, E. Bacry, and A. Arneodo, Int. J. Bifurcation Chaos Appl. Sci. Eng. **4**, 245 (1994).
[12] S. Thurner, M. C. Feurstein, and M. C. Teich, Phys. Rev. Lett. **80**, 1544 (1998).
[13] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik, and H. E. Stanley, Nature (London) **399**, 461 (1999).
[14] A. Marrone, A. D. Polosa, G. Scioscia, S. Stramaglia, and A. Zenzola, Phys. Rev. E **60**, 1088 (1999).
[15] L. A. Nunes Amaral, P. C. Ivanov, N. Aoyagi, I. Hidaka, S. Tomono, A. L. Goldberger, H. E. Stanley, and Y. Yamamoto, Phys. Rev. Lett. **86**, 6026 (2001).
[16] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. -C. Yen, C. -C. Tung, and H. H. Liu, Proc. R. Soc. London, Ser. A **454**, 903 (1998).
[17] T. C. Chiang, H. -C. Yu, and M. -C. Wu (unpublished).
[18] We use the TAQ database published by the New York Stock Exhange to constrcut the intraday returns and Yahoo database (http://finance.yahoo.com/) for the interday returns.
[19] G. M. Mian and C. M. Adam, Appl. Financ. Econ. **11**, 341

(2001).
[20] T. G. Anderson and T. Bollerslev, J. Empirical Finance **4**, 115 (1997).
[21] M. Montero, J. Perelló, and J. Masoliver, Physica A **316**, 539 (2002).
[22] A. A. Dragulescu and V. M. Yakovenko, Quant. Finance **2**, 443 (2002); cond-mat/0203046.
[23] A. C. Silva, R. E. Prange, and V. M. Yakovenko, Physica A **344**, 227 (2004); cond-mat/0401225, and references therein.
[24] S. L. Heston, Rev. Financ. Stud. **6**, 327 (1993).
[25] J. Perelló and J. Masoliver, Phys. Rev. E **67**, 037102 (2003).
[26] J. Perelló, J. Masoliver, and N. Anento, Physica A **344**, 134 (2004).
[27] For example, there are the Ornstein-Uhlenbeck (OU), the Heston, and the exponential OU (expOU) models. See, e.g., Refs. [25,26] and references therein.
[28] N. Iyengar, C. -K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, Am. J. Physiol. **271**, 1078 (1996). Data sets are available from http://physionet.org/physiobank/database/fantasia/, and here we have used data code "f1o06" for illustrations.
[29] M. -C. Wu and C. -K. Hu (unpublished).
[30] Our analysis was based on the daily yen-dollar exchange rates for the New York currency markets from a Wall Street firm over the period October 1994 to December 2003 (111 months). The yen-dollar exchange rates were expressed in European terms (yen/$)—i.e., the foreign currency price of one U.S. dollar.
[31] M. -C. Wu (unpublished).
[32] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).
[33] D. Dhar (unpublished).
[34] M. L. Mehta, *Random Matrices*, 2nd ed. (Academic Press, New York, 1991).
[35] A. M. Sengupta, and P. P. Mitra, Phys. Rev. E **60**, 3389 (1999).
[36] J. D. Noh, Phys. Rev. E **61**, 5981 (2000).
[37] S. Drozdz *et al.*, Physica A **287**, 440 (2000).
[38] N. Vandewalle, F. Brisbois, and X. Tordoir, Quant. Finance **1**, 372 (2001).
[39] G. Bonanno, F. Lillo, and R. N. Mantegna, Quant. Finance **1**, 96 (2001).
[40] R. N. Mantegna, Eur. Phys. J. B **11**, 193 (1999).
[41] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stenly, Proc. Natl. Acad. Sci. U.S.A. **102**, 9424 (2005).