Statistical analysis of the overnight and daytime return

Fengzhong Wang,[1] Shwu-Jane Shieh,[1,2] Shlomo Havlin,[1,3] and H. Eugene Stanley[1]
[1]*Department of Physics and Center for Polymer Studies, Boston University, Boston, Massachusetts 02215, USA*
[2]*Department of International Business, National Cheng-Chi University, Taipei 116, Taiwan, Republic of China*
[3]*Department of Physics and Minerva Center, Bar-Ilan University, Ramat-Gan 52900, Israel*
(Received 4 March 2009; published 21 May 2009)

We investigate the two components of the total daily return (close-to-close), the overnight return (close-to-open), and the daytime return (open-to-close), as well as the corresponding volatilities of the 2215 New York Stock Exchange stocks for the 20 year period from 1988 to 2007. The tail distribution of the volatility, the long-term memory in the sequence, and the cross correlation between different returns are analyzed. Our results suggest that (i) the two component returns and volatilities have features similar to that of the total return and volatility. The tail distribution follows a power law for all volatilities, and long-term correlations exist in the volatility sequences but not in the return sequences. (ii) The daytime return contributes more to the total return. Both the tail distribution and the long-term memory of the daytime volatility are more similar to that of the total volatility, compared to the overnight records. In addition, the cross correlation between the daytime return and the total return is also stronger. (iii) The two component returns tend to be anticorrelated. Moreover, we find that the cross correlations between the three different returns (total, overnight, and daytime) are quite stable over the entire 20 year period.

DOI: 10.1103/PhysRevE.79.056109 PACS number(s): 89.65.Gh, 05.45.Tp, 89.75.Da

I. INTRODUCTION

Financial markets are of great importance for economics and econophysics research [1–21]. A key topic of the market studies is the price dynamics, which could be measured by the price change ("return") and its magnitude ("volatility") [2–20]. Especially, the volatility has important practical implications. For example, it is the key input for option pricing models such as the classic Black-Scholes model and the Cox, Ross, and Rubinstein binomial models [16,17]. Usually financial markets are closed during the night, and all news or events in the night are reflected in the opening price of the next trading day. A day (from former day closing to current day closing) therefore can be decomposed into two sessions, overnight (from former day closing to current day opening) and daytime (from current day opening to closing) sessions. The study of the returns and the volatilities during these two sessions might provide new insights toward better understanding of the financial markets. Practically, this study can help traders to improve trading strategies at the market opening and closing. It can also help investors to analyze the dually traded equities [19].

Recently there were some studies on the returns and volatilities over subday sessions. George and Hwang [18] decomposed the daily return of 200 Japanese stocks and analyzed their volatility patterns. Wang *et al.* [19] studied 15 stocks which are traded in both Hong Kong and London but in different hours. However, there is still lack of a comprehensive analysis of the overnight and daytime price change for a leading market such as the New York Stock Exchange (NYSE). For the daily and high-frequency intraday data, returns and volatilities of stock prices are well studied [2–20]. These studies show that the return and volatility distribution decay as power laws, and the correlations in the returns disappear after a few minutes while the correlations in the volatility time series can exist up to months and even longer [20–25]. It is of interest to examine whether these features persist also in the two component returns and volatilities. Obviously one can assume that the overnight price change behaves statistically different from the daytime change. What are the differences? Furthermore, the influence of the overnight price change on the daytime change is also of interest and should be examined.

In this paper we examine the daily data for all stocks traded in NYSE. First we study the fundamental features of the time series, distribution of the records, and the correlations in the sequence. Three types of functions, power law, exponential, and power law with an exponential cutoff, are tested for the tail of the volatility distribution. We find that the power-law function fits best for most stocks. Then we analyze the long-term memory of each stock using the detrended fluctuation analysis (DFA) method [26–29] and find that the long-term correlations persist in the volatilities of both components. We show that the distribution and the long-term memory of the daytime volatility are more similar to the total volatility, compared to the overnight volatility. Further, we study the cross correlations between the three types of returns (total, overnight, and daytime). The two component returns are found to be weakly anticorrelated but both overnight and daytime returns are strongly correlated with the total return. Interestingly, we find that this behavior is quite stable during the entire 20 year period.

II. DATA AND VARIABLES

We collect the daily opening and closing prices of all securities that are listed in NYSE on December 31, 2007, 2215 stocks in total [30]. The record starts from January 2, 1962, but many stocks have a much shorter history. We do not include the data before the 1987 period for two reasons. First, from 1962 to 1987 there exist only very little data, about 6.5% of all the data points for these 2215 stocks. Sec-

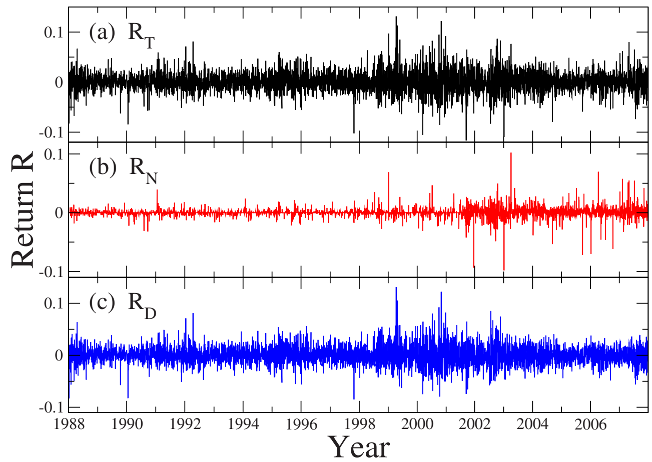

FIG. 1. (Color online) Illustration of the return time series. Three types of return, (a) the total return R_T, (b) the overnight return R_N, and (c) the daytime return R_D of a typical stock, AA, are shown. We can see that the fluctuations of R_N are relatively weaker, and the curve of R_D is more similar to that of R_T.

ond and more important, there was a huge market crash on October 19, 1987 ("Black Monday"), and after that the market was adapted in a great extent. Thus, to reduce the complexity of market structure, we only examine the data from 1988 to 2007, 20 years in total. The length of the 2215 stocks ranges from $N=1000$ to 5000 trading days. Note that many stocks have splits in the 20 year period, which causes a significant change in the price. Therefore, we adjust all prices according to the historical splits. The 2215 stocks cover all industrial sectors, a wide range of the stock market capitalization (from 6×10^6 to 5×10^{11} dollars), and a wide range of the average daily volume (from 500 to 2×10^7 shares a day).

Now we define two basic measures, return R and volatility V. The daily return is the logarithmic change in the successive daily closing prices ("total return"),

$$R_T(t) \equiv \ln(p^{\text{close}}(t)/p^{\text{close}}(t-1)), \qquad (1)$$

the return over the overnight session ("overnight return") is

$$R_N(t) \equiv \ln(p^{\text{open}}(t)/p^{\text{close}}(t-1)), \qquad (2)$$

and the return over the daytime session ("daytime return") is

$$R_D(t) \equiv \ln(p^{\text{close}}(t)/p^{\text{open}}(t)). \qquad (3)$$

Here $p^{\text{close}}(t)$ is the closing price and $p^{\text{open}}(t)$ is the opening price at day t. Note that $R_T(t)=R_N(t)+R_D(t)$, $R_D(t)$, and $R_N(t)$ are in the same day and $R_D(t)$ is after $R_N(t)$. Figure 1 shows the three types of return for a typical stock Alcoa, Inc. (AA) from 1988 to 2007.

The volatility is defined as the absolute value of the return [10–12], i.e.,

$$V(t) \equiv |R(t)|. \qquad (4)$$

Thus, corresponding to the three types of return, we have three types of volatility, the total volatility V_T, the overnight volatility V_N, and the daytime volatility V_D. Note that this volatility definition is not based on any distribution assumption, such as the Gaussian distribution. However, it measures the deviations from the expected rate of return, therefore characterizes the magnitude of price fluctuations and risk of the asset.

TABLE I. Number of good fits of the volatility tail distribution for the 2215 NYSE stocks. Good fits refer to the cases where the null hypothesis is not ruled out for 1% significance level.

Volatility V	V_T	V_N	V_D
Power law	2066	1868	2066
Exponential	1693	644	1756
Power law with cutoff	1755	1772	1728

III. TAIL OF VOLATILITY DISTRIBUTION

The tail distribution accounts for large fluctuations and events which are very important for risk analysis. By definition [Eq. (4)], the volatility aggregates both positive and negative returns and has better statistics. In addition, the distribution of the return is approximately symmetric in the two tails [20]. Therefore we focus on the tail distribution of the volatility. As a stylized fact of econophysics research, the cumulative distribution function (CDF) [31] of volatilities has a "fat tail" which is usually characterized by a power law [20–25],

$$P(x) \sim x^{-\zeta}, \qquad (5)$$

where ζ is the tail exponent. A classical approach to fit the tail is using the maximum likelihood estimator, which is called Hill estimator for a power-law tail [24,25,32]. The goodness of fit is tested by the Kolmogorov-Smirnov (KS) statistic D [33,34], the maximum absolute difference between the cumulative distribution of the measured distribution $P(x)$ and that of the fit $S(x)$, i.e.,

$$D \equiv \max(|P(x) - S(x)|), \qquad (6)$$

for all volatility values in the tail [35]. When D is larger than a certain value, which is called critical value (CV), the null hypothesis that the distribution follows a power law is rejected. The CV is determined by the significance level and data size N. In this paper we choose a significance level of 1% and the corresponding $\text{CV}=1.63/\sqrt{N}$.

To further test the volatility tail, we also try two other distribution functions in the same range and using the same method. One is the exponential distribution function,

$$P(x) \sim e^{-x/x^*}, \qquad (7)$$

where x^* is a characteristic scale. The other is a power-law function with an exponential cutoff,

$$P(x) \sim x^{-\zeta} e^{-x/x^*}. \qquad (8)$$

We examine the tail distribution of V_T, V_N, and V_D for the 2215 NYSE stocks. The number of fits that the null hypothesis was valid under 1% significant level ("good fit") is listed in Table I. For the power-law distribution, only a small portion (10%) of the three types of volatilities are ruled out,

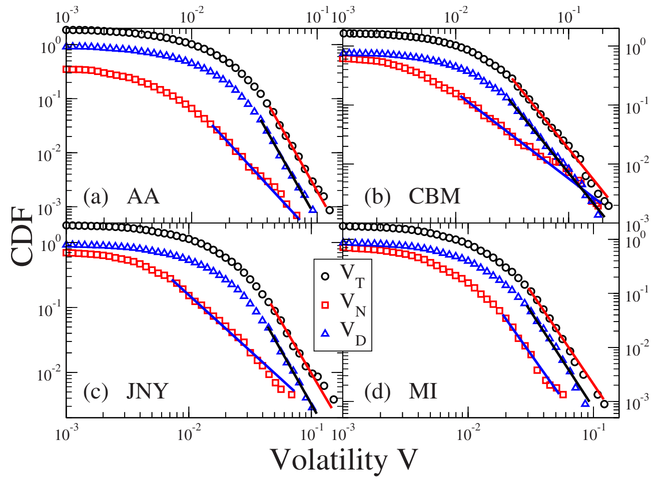

FIG. 2. (Color online) Typical cumulative distribution of the volatilities and power-law fit to the tails. For four typical stocks, (a) AA, (b) CBM, (c) JNY, and (d) MI, three types of volatility, total volatility V_T (circles), overnight volatility V_N (squares), and daytime volatility V_D (triangles) are demonstrated. The solid lines are power-law fits to the distribution tails. Note that the curves for V_T (circles) almost coincide with those of V_D, and thus they are vertically shifted for better visibility.

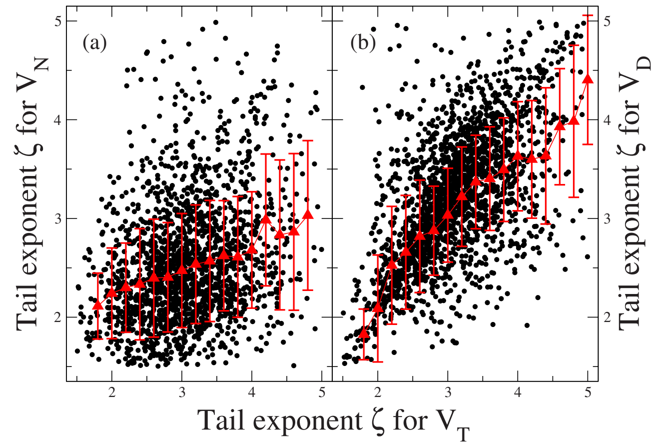

FIG. 3. (Color online) Relation between the tail exponent ζ for the total volatility V_T and that for the two component volatilities, (a) the overnight volatility V_N, and (b) the daytime volatility V_D. A point represents a stock which has good power-law fit to the tail for the corresponding two types of volatilities. 1812 out of the 2215 NYSE stocks are exhibited in panel (a) and 2001 stocks are exhibited in panel (b). To show the tendency, we divide the entire data set into equal-width subsets according the value of ζ for V_T and calculate the mean values and standard deviations in these subsets, as shown by the triangles and the error bars, respectively. Both cases clearly show tendencies but that for the daytime volatility is stronger, indicating that V_T is more influenced by V_D. Moreover, ζ for all three types of volatilities are distributed in a relatively narrow range and centered around 3.

which manifests that the tail is well characterized by the power-law function for the broad market. For the exponential hypothesis, almost half (38%) of all the cases are ruled out. Moreover, about 98% out of the good exponential fits, the power-law hypothesis is not ruled out either. As a whole, the exponential function is poor for characterizing the tail, compared to the power-law function. For the power law with an exponential cutoff, the percentage of good fits is 79% over the three volatilities, which is slightly lower than that for the power law. Besides, 99% out of them do not reject the power-law hypothesis either. Therefore, we conclude that the power law is the best among the three distributions.

In Fig. 2, we plot the CDF of V_T, V_N, and V_D for four typical stocks, namely, AA, Cambrex Corp. (CBM), Jones Apparel Group, Inc. (JNY), and Marshall and Ilsley Corp. (MI). These stocks belong to diverse industrial sectors and their capitalization vary in a wide range, from 27 billion dollars for AA to 0.25 billion dollars for MI. As seen in Fig. 2, the tails are well fitted by power laws. Interestingly, the tails of V_D almost always decay faster than the tails of V_N, and V_T lies between the two component volatilities. Moreover, the log-log slope (tail exponent ζ) of V_T is closer to that of V_D, indicating the daytime return contributes more to the total return. To test this finding for the broad market, we plot in Fig. 3 the relation between the tail exponent ζ of V_T and ζ of the two component volatilities for the 2215 stocks. Both scatter plots show a certain dependence (as shown by the solid curves, which are averages over different bins of ζ of V_T), but the correlation between V_T and V_D is obviously stronger, which is consistent with Fig. 2. For all the three types of volatilities, ζ is distributed in a certain range from 1.5 to 5 and centered around 3. The averages of ζ are as follows: $\langle\zeta\rangle \approx 2.6$ for V_N is lower than $\langle\zeta\rangle \approx 3.2$ for V_D, while $\langle\zeta\rangle \approx 3.1$ for V_T is between the two component volatilities and it is slightly smaller than that for V_D. In this paper $\langle\cdot\rangle$ stands for the average over the data set. This behavior suggests that the daytime return influences the total return more than the overnight return.

IV. CORRELATIONS IN RETURNS AND VOLATILITIES

After analyzing the distribution, a question naturally arises: how are these values organized in the time sequence? For the investors, the temporal structure is of special interest because it determines how and when to trade. The time organization in a time series can be characterized by the two-point correlation. It is known that the total return only has short-term correlations and the total volatility has long-term correlations [20–25]. Now we examine the correlations in each of their two components (overnight and daytime).

It is well known that financial time series are usually nonstationary. In such cases, the conventional methods for correlations such as autocorrelation and spectral analysis have spurious effects. To avoid the artifact correlations arising from nonstationarity, we employ the DFA method, which is based on the idea that a correlated time series could be mapped to a self-similar process by integration, and removing systematically trends in order to detect the long-term correlations in the time series [26–29]. After removing polynomial trends in every equal-size box of ℓ points, DFA computes the root-mean-square fluctuation $F(\ell)$ of a time series and determine the correlation exponent α from the scaling relation

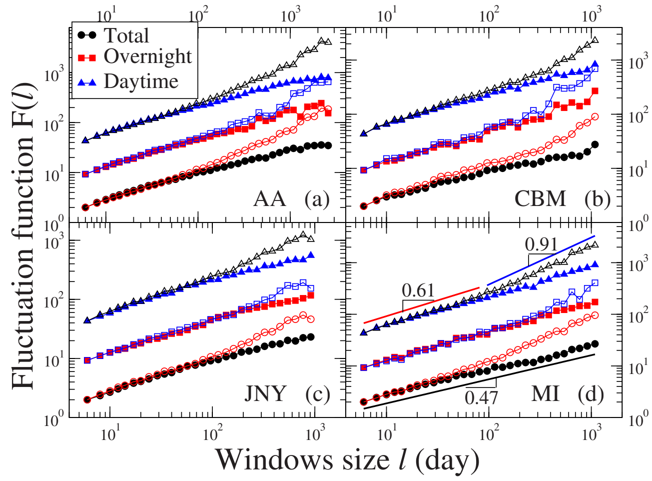

FIG. 4. (Color online) DFA fluctuation function F vs window size ℓ for the returns (filled symbols) and the volatilities (open symbols). The four panels are for stocks AA, CBM, JNY, and MI, respectively. For each case, three types of data, total (circles), overnight (squares), and daytime (triangles), are shown. Note that the curves are vertically shifted for better visibility. To obtain the correlation exponent α, we fit all the curves with power laws, as illustrated by the solid lines in panel (d). For the volatility, the exponent α is significantly different for short and long time scales; thus we split the entire range into two regimes and fit them separately.

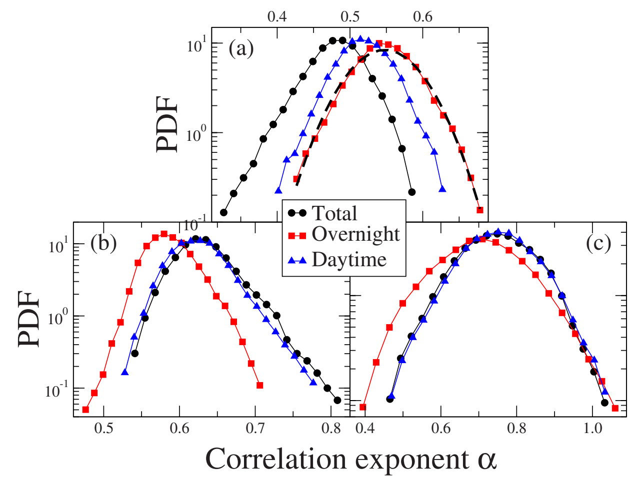

FIG. 5. (Color online) Distribution of the correlation exponent α for (a) the returns, (b) the volatilities of the short time scales, and (c) the volatilities of the long time scales. Three types of returns and volatilities, total (circles), overnight (squares), and daytime (triangles), are shown. All distributions approximately follow the normal distribution. For example, a normal distribution fit on the overnight return is shown by the dashed line in panel (a). For the volatilities, the curves for the total and the daytime almost collapse into a single curve in panels (b) and (c), suggesting that the total return is more influenced by the daytime return.

$$F(\ell) \sim \ell^{\alpha}, \tag{9}$$

where the exponent $\alpha \in (0,1)$, called correlation exponent, characterizes the autocorrelation in the sequence. It is uncorrelated if $\alpha = 0.5$, positively correlated if $\alpha > 0.5$, and anticorrelated if $\alpha < 0.5$. In Fig. 4, we plot the DFA curves for the returns and volatilities of the total, overnight, and daytime sequences for four typical stocks. The values of α are obtained by the power-law fit to the fluctuation function, as illustrated by the solid lines in Fig. 4(d). For all three types of returns, α is close to 0.5, and therefore there are no long-term correlations. For the volatilities, the fluctuation function is more complicated. The slopes (in log-log scale) of different regions are significantly different. Thus, we divide the whole curve into two equal-size regions in the logarithmic scale and fit them separately, as shown by the solid lines in Fig. 4(d).

To test the universality of our findings, we plot in Fig. 5 the probability density function (PDF, which is the derivative of CDF) of α for the three returns as well as for the short and long time scales of the volatilities. For the returns [Fig. 5(a)], the distributions are centered around 0.5, $\alpha = 0.48 \pm 0.04$ for the total, $\alpha = 0.55 \pm 0.05$ for the overnight, and $\alpha = 0.52 \pm 0.04$ for the daytime. Here and in the following, the error bars are the standard deviations over the 2215 stocks. These error bars are quite small representing quite narrow distributions. This result is consistent with earlier studies, where no long-term correlations were found for the returns [20]. For the volatilities at short time scales [Fig. 5(b)], the distributions are centered around 0.6, $\alpha = 0.63 \pm 0.04$ for the total [20], $\alpha = 0.59 \pm 0.03$ for the overnight and $\alpha = 0.63 \pm 0.04$ for the daytime. For the volatilities at long time scales [Fig. 5(c)], $\alpha = 0.75 \pm 0.10$ for the total [20], $\alpha = 0.71 \pm 0.12$ for the overnight, and $\alpha = 0.75 \pm 0.10$ for the daytime. For all time scales, the volatility α values are significantly larger than 0.5, suggesting long-term correlations in the volatility sequences. In addition, the α values of the long-term scales are systematically larger than that of the short-term scales. This multiscaling behavior indicates that the correlation becomes stronger for longer times. Moreover, all distributions are relatively narrow for both returns and volatilities, suggesting a universal feature over the entire market. We also see that the curves of the total and daytime almost collapse onto a single curve, while the curve of the overnight departs away from them, supporting again that the daytime return contributes more than the overnight return to the total return.

Now we address the question if there is a relation between the correlation exponents α of the two components of the return and volatility. If a certain stock has large (small) α for one component, does it also have large (small) α in the other component or in the total? To test this, we employ the cross-correlation function to quantitatively compare them. The cross correlation (also called the Pearson coefficient) between variables x and y is

$$C(x,y) \equiv \frac{\langle xy\rangle - \langle x\rangle\langle y\rangle}{\sigma(x)\sigma(y)}. \tag{10}$$

Here σ stands for the standard deviation, i.e., for variable x, $\sigma(x) \equiv \sqrt{\langle x^2\rangle - \langle x\rangle^2}$. For our case, x and y are vectors representing the three sequences of α (total, overnight, daytime) for the return, short-time, and long-time volatilities of all companies. The companies are in the same order for all se-

TABLE II. Cross correlation between the α values of the three types of returns and volatilities for the 2215 NYSE stocks. We divide the 2215 stocks into ten equal-size subsets and calculate the cross correlation for every subset. The error bar is the corresponding standard deviation of the ten cross correlations. The value in the parenthesis is the corresponding cross correlation between two shuffled α records.

Cross correlation C	C (total, overnight)	C (total, daytime)	C (overnight, daytime)
Return	0.25 ± 0.07	0.51 ± 0.08	0.48 ± 0.08
	(−0.00)	(−0.03)	(0.02)
Volatility (short time scales)	0.29 ± 0.09	0.80 ± 0.04	0.24 ± 0.04
	(0.04)	(−0.02)	(0.03)
Volatility (long time scales)	0.56 ± 0.06	0.90 ± 0.02	0.52 ± 0.05
	(0.01)	(0.02)	(−0.02)

quences. As shown in Table II, all cross correlations are significantly larger than that of the shuffled records (values in the parenthesis), suggesting strong relations between the different returns and volatilities. Note again that the total-daytime pair is always the strongest one, which is in agreement with the assumption that the total return and volatility are significantly more influenced by the daytime return and volatility than by the overnight return and volatility.

V. RELATION AMONG TOTAL, OVERNIGHT, AND DAYTIME RETURNS

The overnight return and the daytime return are the price changes over different sessions of a trading day, and they make the total return. It is interesting to examine now if the three returns of the same stock are cross correlated. This will test the question, e.g., how the changes in the day time are related to those of the night time or the total. The cross-correlation function [Eq. (10)] examines the two time series without any time lag. However, there might be some time delays between the two time series, and therefore we shift the two sequences by time lag Δt to test this possibility. Moreover, the comparison between the cross correlations with different lags allows us to examine the significance of a cross-correlation value. Therefore, we use the generalized cross correlation with the time lag Δt, i.e.,

$$C_{\Delta t}(x,y) \equiv \frac{\langle x(t)y(t+\Delta t)\rangle - \langle x\rangle\langle y\rangle}{\sigma(x)\sigma(y)}, \tag{11}$$

between the two time series $x(t)$ and $y(t)$. Note that Eq. (10) is the special case of Eq. (11) with $\Delta t=0$. In general one tests the position of the maximum (minimum if it is anticorrelated) of $C_{\Delta t}$ which may occur at $\Delta t=\tau$ and τ is called the time delay [36]. Here we find that the maximum of $C_{\Delta t}$ is always for $\Delta t=0$ (as shown in Fig. 6).

In this paper we use $C_{\Delta t}$ to test the significance of the cross correlation at $\Delta t=0$. If $C_{\Delta t=0}$ is significantly different (higher or lower) from $C_{\Delta t\neq 0}$, the cross correlation can be regarded as reliable. Quantitatively, we use the standard deviation of $C_{\Delta t\neq 0}$ values over the range $-20 \le \Delta t \le 20$, $\sigma(C_{\Delta t\neq 0})$, to test the reliability of the cross correlation [36]. As examples, we plot in Fig. 6 the cross correlations of the three pairs of returns for the four typical stocks, AA, CBM, JNY, and MI (other stocks have similar features). For both $C(R_T,R_N)$ and $C(R_T,R_D)$, the cross correlations at $\Delta t=0$ are more than ten times higher than their $\sigma(C_{\Delta t\neq 0})$ so they are very robust. However, for $C(R_N,R_D)$, the cross correlations vary with the stock. Some of them have significant cross-correlation values but some of them are in the range of their $\sigma(C_{\Delta t\neq 0})$. Since R_N and R_D covers different periods, there could be some strong correlations or almost independent; it is reasonable that the cross correlation varies in a wide range. On the other hand, R_T always shares a part of the changes with its two component returns and deduce strong positive cross correlations.

Next we examine the three pairs of cross correlations $C_{\Delta t=0}$ for all the 2215 stocks (in the following, the function C refers to $C_{\Delta t=0}$ if the Δt subscript is missing). Their distributions are plotted in Fig. 7. For each pair, the cross correlations are distributed in a certain range. The cross correlation between the total return and the daytime return, $C(R_T,R_D)$

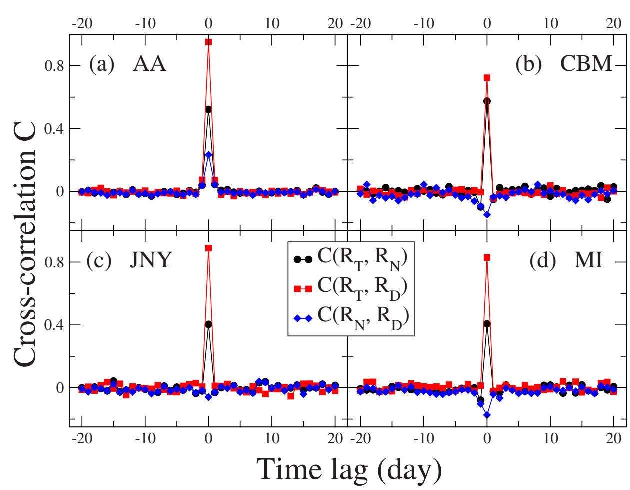

FIG. 6. (Color online) Reliability of the cross correlations between the three types of returns, the total return R_T, the overnight return R_N, and the daytime return R_D. For all four stocks (a) AA, (b) CBM, (c) JNY, and (d) MI, the two cross correlations with respect to the total return are significantly larger than their cross correlations with the time lags, which suggests that both component returns are strongly positively correlated with the total return. However, the cross correlation between the two component returns varies with the stocks, e.g., it is positive for AA and negative for MI.

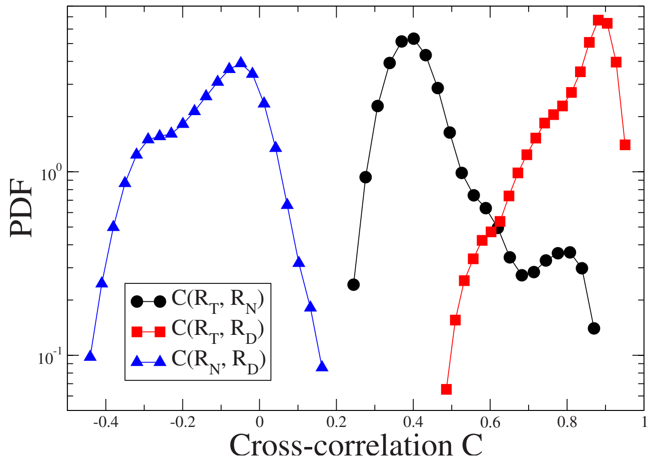

FIG. 7. (Color online) Distribution of the cross correlations C among the returns R_T, R_N, and R_D for the 2215 NYSE stocks. Both cross correlations with respect to the total return are significantly larger than 0 and that for the daytime return is stronger, suggesting that the total return is more correlated with the daytime return. The cross correlation between the two component returns is relatively more distributed toward negative values, indicating that the two component returns tend to be anticorrelated.

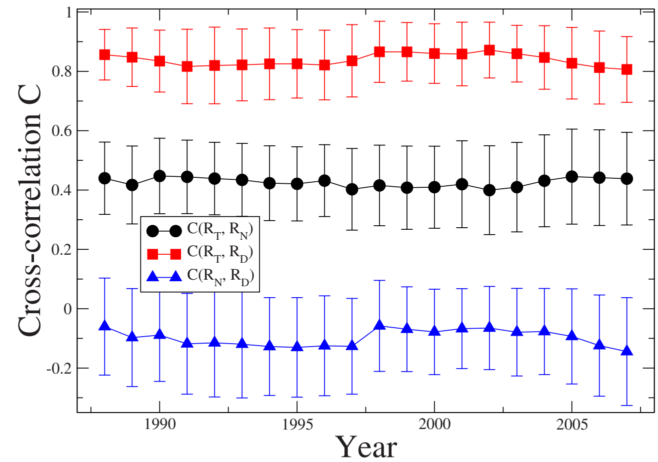

FIG. 8. (Color online) Evolution of the cross correlations between the three returns, R_T, R_N, and R_D, from 1988 to 2007. Here a point represents the average over the cross correlations of the 2215 NYSE stocks in a 1 year period, and the error bar is the corresponding standard deviation. Clearly, there are no significant changes for the cross correlations over the 20 year period studied.

$=0.8\pm0.1$ (mean value and standard deviation over the 2215 stocks), is always the largest value in the three pairs. The cross correlation between the total and the overnight, $C(R_T,R_N)=0.4\pm0.1$, is still high but significantly smaller than the $C(R_T,R_D)$ values. The cross correlation between day and night, $C(R_N,R_D)=-0.1\pm0.1$, is distributed around 0 with more tendency to have negative values. In summary, the total return is more synchronized with the daytime return. It is also interesting to note that there are significantly more stocks that have negative correlations between R_N and R_D. For example, 567 out of the 2215 stocks have values of $C(R_N,R_D)<-0.2$. This implies that the probability is relatively high for a large positive (negative) overnight return to be followed by a large negative (positive) daytime return. The overnight return and the daytime return tend to be slightly anticorrelated, and the total return usually moves in the same direction as the daytime return.

Due to many factors, such as changes in the regulations or new technologies, the market evolves with time. An interesting question arises: is the cross correlation stable in the sample years studied? To test the stability of the cross correlations, we recalculate the cross correlations year by year. The records in 1 year are enough to calculate the cross correlation and more importantly, the equity market in such a short period can be assumed stable. In Fig. 8 we plot the averages and standard deviations (as error bars) of the cross correlations over the 2215 stocks against the year. For the three types of cross correlations, the curves only slightly vary with the years and all changes lie within the error bars. Moreover, the error bars are almost the same for all years, which clearly shows that the cross correlation is quite stable over the 20 year period studied.

VI. DISCUSSION

Returns and volatilities might be affected by some factors, such as the market capitalization and the mean volume [12]. To test this for the entire stock market, we investigate the relation between the factors, including the capitalization and mean volume, and the measures, such as the tail exponent ζ, the correlation exponent α, and the cross correlations between the three returns. There are some tendencies between these factors and measures. However, most of these tendencies are in the range of the error bars, which suggests no significant dependence between the two factors and three measurements. The behavior of the three measures is quite universal over the entire market. To better understand the complexity of the equity market, the connection between different measurements and factors of stocks might need to be further analyzed.

In summary, we examined the distributions of the total, overnight, and daytime volatilities. Compared to the exponential and power law with an exponential cutoff, power-law distribution is found to be mostly better. The tail exponent ζ is distributed among the different stocks between 1.5 and 5 for the three types of volatility. We also analyzed the correlations in returns and volatilities of the components using the DFA method. For both returns, there are no long-term correlations. However, for both volatilities, there are long-term correlations in all time scales and the correlations are even stronger in the long time scales. For the tail distribution and for the long-term correlations, the results of the two component returns and volatilities are similar to the total return and volatility. Moreover, the records of the daytime are more similar to the total of the same stock, suggesting that the daytime return contributes more to the total return. To better compare these similarities, we also studied the cross correlations between the different types of return and found consistent behaviors, i.e., the daytime is more correlated with the total compared to the night time. Further, the cross correlation between the overnight return and the daytime return varies for different stocks, and interestingly, a significant fraction of the 2215 stocks is far below 0. This finding suggests that the daytime return has a considerable probability to strongly anticorrelate with the overnight return. Furthermore, we examined the cross correlations year by year and found that the behavior is quite stable over the 20 year period.

ACKNOWLEDGMENTS

We thank K. Yamasaki for helpful discussions, the National Science Foundation and Merck Foundation for financial support. S.-J.S. thanks the NSF of Taiwan for financial support.

[1] *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes*, edited by A. Bunde, J. Kropp, and H. J. Schellnhuber (Springer-Verlag, Berlin, 2002).
[2] R. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, England, 2000).
[3] *Econophysics: An Emerging Science*, edited by I. Kondor and J. Kertész (Kluwer, Dordrecht, 1999).
[4] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, England, 2003).
[5] N. F. Johnson, P. Jefferies, and P. M. Hui, *Financial Market Complexity* (Oxford University Press, New York, 2003).
[6] H. Takayasu, H. Miura, T. Hirabayashi, and K. Hamada, Physica A **184**, 127 (1992); H. Takayasu, A. H. Sato, and M. Takayasu, Phys. Rev. Lett. **79**, 966 (1997); H. Takayasu and K. Okuyama, Fractals **6**, 67 (1998).
[7] Z. Ding, C. W. J. Granger, and R. F. Engle, J. Empirical Finance **1**, 83 (1993).
[8] R. A. Wood, T. H. McInish, and J. K. Ord, J. Financ. **40**, 723 (1985).
[9] L. Harris, J. Financ. Econ. **16**, 99 (1986).
[10] K. Yamasaki, L. Muchnik, S. Havlin, A. Bunde, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **102**, 9424 (2005).
[11] F. Wang, K. Yamasaki, S. Havlin, and H. E. Stanley, Phys. Rev. E **73**, 026117 (2006); F. Wang, P. Weber, K. Yamasaki, S. Havlin, and H. E. Stanley, Eur. Phys. J. B **55**, 123 (2007); W.-S. Jung, F. Z. Wang, S. Havlin, T. Kaizoji, H.-T. Moon, and H. E. Stanley, *ibid.* **62**, 113 (2008); I. Vodenska-Chitkushev, F. Z. Wang, P. Weber, K. Yamasaki, S. Havlin, and H. E. Stanley, *ibid.* **61**, 217 (2008).
[12] F. Wang, K. Yamasaki, S. Havlin, and H. E. Stanley, Phys. Rev. E **77**, 016109 (2008); F. Wang, K. Yamasaki, S. Havlin, and H. E. Stanley, *ibid.* **79**, 016103 (2009).
[13] P. Weber, F. Wang, I. Vodenska-Chitkushev, S. Havlin, and H. E. Stanley, Phys. Rev. E **76**, 016109 (2007).
[14] Z. Eisler, I. Bartos, and J. Kertész, Adv. Phys. **57**, 89 (2008).
[15] W.-S. Jung, O. Kwon, F. Wang, T. Kaizoji, H.-T. Moon, and H. E. Stanley, Physica A **387**, 537 (2008).
[16] F. Black and M. Scholes, J. Polit. Econ. **81**, 637 (1973).
[17] J. C. Cox and S. A. Ross, J. Financ. Econ. **3**, 145 (1976); J. C. Cox, S. A. Ross, and M. Rubinstein, *ibid.* **7**, 229 (1979).
[18] T. J. George and C. H. Hwang, J. Financ. Quant. Anal. **30**, 313 (1995).
[19] S. S. Wang, O. M. Rui, and M. Firth, J. Int. Money Finance **21**, 265 (2002).
[20] Y. Liu, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, Physica A **245**, 437 (1997); P. Cizeau, Y. Liu, M. Meyer, C.-K. Peng, and H. E. Stanley, *ibid.* **245**, 441 (1997); Y. Liu, P. Gopikrishnan, P. Cizeau, M. Meyer, C.-K. Peng, and H. E. Stanley, Phys. Rev. E **60**, 1390 (1999); V. Plerou, P. Gopikrishnan, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley, *ibid.* **60**, 6519 (1999); V. Plerou, P. Gopikrishnan, X. Gabaix, L. A. N. Amaral, and H. E. Stanley, Quant. Finance **1**, 262 (2001); V. Plerou, P. Gopikrishnan, and H. E. Stanley, Phys. Rev. E **71**, 046131 (2005).
[21] B. B. Mandelbrot, J. Bus. **36**, 394 (1963).
[22] T. Lux, Appl. Financ. Econ. **6**, 463 (1996).
[23] U. A. Muller, M. M. Dacorogna, and O. V. Pictet, in *A Practical Guide to Heavy Tails*, edited by R. J. Adler, R. E. Feldman, and M. S. Taqqu (Birkhäuser, Boston, 1998), p. 83.
[24] V. Plerou and H. E. Stanley, Phys. Rev. E **76**, 046109 (2007); **77**, 037101 (2008).
[25] A. Clauset, C. R. Shalizi, and M. E. J. Newman, e-print arXiv:0706.1062.
[26] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Phys. Rev. E **49**, 1685 (1994).
[27] A. Bunde, S. Havlin, J. W. Kantelhardt, T. Penzel, J.-H. Peter, and K. Voigt, Phys. Rev. Lett. **85**, 3736 (2000).
[28] K. Hu, P. Ch. Ivanov, Z. Chen, P. Carpena, and H. E. Stanley, Phys. Rev. E **64**, 011114 (2001); L. Xu, P. Ch. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley, *ibid.* **71**, 051101 (2005); Z. Chen, K. Hu, P. Carpena, P. Bernaola-Galvan, H. E. Stanley, and P. Ch. Ivanov, *ibid.* **71**, 011104 (2005).
[29] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde, Physica A **295**, 441 (2001).
[30] Historical stock data is available at http://finance.yahoo.com. To obtain good statistics, we only choose the stocks with more than 1000 records.
[31] In this paper the cumulative distribution function is often called the complementary cumulative distribution function, $P(x) \equiv P(x' > x)$, and P is the probability of variable x. For simplicity we call it "cumulative distribution" or CDF.
[32] B. M. Hill, Ann. Stat. **3**, 1163 (1975).
[33] M. A. Stephens, J. Am. Stat. Assoc. **69**, 730 (1974).
[34] R. Engle and J. Russel, Econometrica **66**, 1127 (1998).
[35] The range of fit for the Hill estimator is not fixed [24,25]. Here we shift the lower bound of the range, examine the corresponding KS statistics, and choose one that has the minimum D value, as suggested in Ref. [25]. To obtain good statistics, we use the highest 3% of all data points as the minimum range. For the three types of volatilities of the 2215 stocks, the average number of data points in the tail is about 430 (the tail averagely covers the highest 10% data points). In order to put all fits on the same footing, we use the same range for a stock in the two other types of fit, the exponential and the power law with exponential cutoff.
[36] K. Yamasaki, A. Gozolchiani, and S. Havlin, Phys. Rev. Lett. **100**, 228501 (2008).