

# Mathematical methods for modelling price fluctuations of financial time series

Sabyasachi Guharay  
Operations Research & Financial Engineering  
Princeton University

August 8, 2002

## 1 Abstract

Statistical analysis of financial time series is studied. We use wavelet analysis to study signal to noise ratios along with auto-correlation function to study correlation length for time series data of daily stock prices for specific sectors of the market. We study the "high beta" stocks versus the "low beta" stocks. We sample ten companies from both of these sectors. We find that the signal to noise ratio is not uniformly high for the "high beta" classified stocks nor is the correlation length large for the "high beta" classified stocks. We explain reasons for this and possible further applications.

## 2 Sponsors

This work was sponsored at the Indiana University Mathematics REU in the summer of 2001. This REU is headed by Professor Allan Edmonds (edmonds@indiana.edu). The advisor for this work was Professor Joseph Stampfli (stampfli@indiana.edu).

## 3 Keywords List and AMS Classification

Correlations, financial mathematics, wavelet applications, stock sectors analysis, quantitative finance. AMS Classifications: 91B84, 62P20, 62P05.

## 4 Introduction

Statistical analysis of time series has been a problem of considerable recent interest. With the surge of data outpouring from various fields such as biology, geophysics, finance (Human Genome Project, digitization of fingerprint data, seismic data etc.) , it is becoming imperative to develop and use proper mathematical tools for classification and understanding these systems. For example, in the field of DNA sequence analysis, there exists immense mathematical literature [1-5] . More recently (as of late 1995 and beyond), Wall Street analysts have started using calculus based methods for their own analysis. These advances have motivated certain groups of physicists and applied mathematicians to try to apply some of their own classical methodologies to understanding dynamics of stock market fluctuations. Besides the classical Black-Scholes formulation [6], a tremendous amount work using wavelet applications [7-9], Lévy distributions and spectral analysis [10,11], multifractal models [12-14] (to name a few) have been used as "tools" to understand the mathematical signature in these financial time series. The basic goal of all these methodologies is to understand the mathematical properties of "long-term" memory versus "short term" memory (long range versus short range correlations)of the market.

In general, most of this research work [7-14], has been applied to measuring the fluctuations of market indices. For example, a lot of work has been done in efforts to detect trends in the S&P 500 (for various different time periods). There has also been some work done on currency exchange dynamics [15]. In most of these cases, the correlation content has been measured for long-term periods (lengths of at least five years). There seems to be a dearth of comprehensive work on looking at differences in actual stock prices (that of various companies). This paper will deal with stock prices fluctuation analysis for certain companies. We have picked these companies from different sectors of the market. Our goal is to see if one can detect similarities or differences (mathematically speaking) between different sectors of the market in the a time frame of one year.

This paper will be organized in the following manner: motivation is narrated next, then we will talk about the data sets we have chosen, next the mathematical methodologies will be described, afterwards the results will be presented, then we will have discussions on the results and conclusions and finally we will end with talking about future plans for this research.

## 5 Motivation

The distribution of price fluctuations are important from a theoretical point of view, are helpful in understanding market dynamics, and in pricing derivative products [17]. The observed complex dynamics in the fluctuations shows an indication of the agent interactions and overall organization of the market. Similarities or differences in the trends for different sectors can give insight as to the relative "stability" of that sector. For example, it is well understood in finance that there are stable sectors in the market such as food processing. Regardless of the status of the economy, companies such as Heinz should have relatively little volatility. However, regarding some newly developing sectors (such as bioinformatics) it is difficult define a relative numerical measure of its "stability". Thus, it could be very useful to develop a quantitative relative measure of the "stability" of certain sectors. We will investigate into these property from a price and a return point of view.

## 6 Data sets

We obtained data for twenty different companies. We obtained ten different companies from the "dogs of the dow" sector. This classification is used by MSNBC (*moneycentral.msn.com*). This term generally refers to stocks which have shown a high rate of return and low volatility for a particular time period. Similarly, there is a classification for the stocks which have in general a high volatility. Throughout this paper, they will be referred to as "high beta stocks". Like before, ten different companies from this sector has been studied. The companies studied from the "dogs of the dow" sector are the following:

- Merck Corp.
- American Express
- Coca-Cola Corp.
- Dell Corp.
- Eastman Kodak Corp.
- General Electric Corp.

- Home Depot Corp.
- Intel Corp.
- McDonald's Corp.
- Exxon Corp.

The companies studied from the "high beta stocks stocks" sector are the following:

- Analog Devices Inc.
- Electricidade de Portugal, S.A.
- E-Trade Inc.
- Goldman Sachs
- Jabil Circuit Inc.
- Lehman Brothers Holdings Inc.
- McMoran Exploration Corp.
- Sprint - PCS Corp.
- T D Waterhouse Group Inc.
- Crown Castle International Corp.

All of these data sets were downloaded from Wharton Business School[16]. The daily closing prices were extracted from here for the time period between January of 2000 till the end of December 2000. This generates a total of 251 data points for each company. Since some of the companies could theoretically have stock splits within a year's time frame, we computed a factor to adjust prices, *facpr*, for all the stocks selected. This factor is used to adjust stock prices after a distribution so that a comparison can be made on an equivalent basis between prices before and after the distribution[16]. Here there are three types of distributions.

- In the case of cash dividends, we set the *facpr* to 0.

- In the case that there are mergers, liquidations, or cases where the total security disappears, the convention argues to set  $facpr$  to -1.
- For stock splits,  $facpr$  is defined as the number of additional shares per old share issued:

$$facpr = \left( \frac{s(t) - s(t')}{s(t')} \right) \quad (1)$$

where  $s(t)$  represents the number of shares outstanding,  $t$  is a date on or after the exact date of the split, and  $t'$  is a date before the split [wrds.whaton].

This gives us the actual data set. However, for our analysis, we will have to manipulate this data slightly from the present state. We wish to analyze two separate events: distribution of daily returns and the distribution of prices. The classical definition of the stock price distribution is the following:

$$S(t) = S_0 e^{\mu t + \sigma W_t} \quad (2)$$

where  $S_0$  is the initial price of the stock,  $\mu$  is drift or average growth rate,  $t$  is the time,  $\sigma$  is the volatility, and  $W_t$  represents the brownian motion drift term which is normally distributed with mean 0 and standard deviation of  $\sqrt{t}$ .

So, defining the daily return  $r$  as  $\frac{S(t+\Delta t)}{S(t)}$  (where  $\Delta t = 1$  day) it is evident that after taking logarithms in equation 2 one obtains the following:

$$\log \frac{S(t + \Delta t)}{S(t)} \propto \mu \Delta t \quad (3)$$

Thus, for computational simplicity, we define the return  $r$  as the following:

$$r = \log \frac{S(t + \Delta t)}{S(t)} \quad (4)$$

Now we will address the issue for stock prices. We obtained all the daily closing stock prices for each of the aforementioned companies. This was done for a period of one year. These prices are all normalized using equation (1). Let us call these prices  $S(t)$  ( $1 \leq t \leq 251$ ). Since certain companies have stocks that are valued very high (such as yahoo.com was valued greater than \$400 per share during a period) while other companies are very low (less than a dollar per share), we wanted to have a uniform range for all the stock prices. So we did the following to provide a uniform range:

- For all values of  $t$  for a particular company, we determine  $\alpha = \max S(t)$
- Next, for all values of  $t$ , we define  $\widehat{S}(t) = \frac{100 * S(t)}{\alpha}$

The above steps ensure that all stock prices are distributed between  $(0, 100]$ . We perform one more scaling on  $\widehat{S}(t)$ . For simplicities in the some future calculations (elaborated in Sec. 7.1), we will require the following final condition on the stock prices:

$$\sum_{t=1}^n \widehat{S}(t) = 0 \tag{5}$$

Thus we subtract the average value of all the prices from each price to ensure that equation (5) holds. Let's call these final prices as  $\widetilde{S}(t)$ . In fig. 1, we show the distribution of prices for Dell company. In this figure we are plotting the  $S(t)$  versus  $t$ . In fig. 2 we show the distribution of normalized and scaled prices of Dell company. Here we are plotting  $\widetilde{S}(t)$  versus  $t$ . Notice that there are negative values for  $\widetilde{S}(t)$ . However, the trend for the fluctuation is preserved.

## 7 Mathematical Methods

We use two mathematical methods, namely, wavelet approximation and auto-correlation function analysis on the stock data sets. Using the wavelet approximation, we compute the signal to noise ratio (*snr*). We use the auto-correlation function to compute the correlation length. We will describe each of these methodologies in upcoming subsections.

### 7.1 Wavelet Approximation (Analytical case)

We will begin by defining a wavelet. Let  $f$  be a function with the following property:

$$\int_{-\infty}^{\infty} |f(x)|^2 dx < \infty$$

The above is the same as saying the following:

$$f \in L^2(\mathbf{R})$$

Suppose  $\psi$  is a real valued function which is supported on  $[0,1]$ . Let us define the following:

$$\psi_{k,n}(x) = 2^{k/2}\psi(2^kx - n) \quad n, k \in \mathbf{Z} \quad (6)$$

Then there is a wavelet basis which is  $(\psi_{k,n})$ . From the theoretical viewpoint [18], the function  $\psi$  is called a wavelet if it meets the following three conditions:

- One can use  $(f, \psi_{k,n})$  to determine the original function  $f$  (here  $(f, \psi_{k,n})$  is the usual  $L^2$  inner product).
- The  $(f, \psi_{k,n})$  forms an orthonormal basis.

The simplest case for  $\psi$  is called the Daubechies 1 wavelet [18]. This is also called the Haar wavelet (the term which we will use throughout this paper). The Haar wavelet is defined as the following:

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \leq x \leq 0.5 \\ -1 & \text{for } 0.5 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Along with satisfying equation (7) there is one final condition for the Haar wavelet:

$$\int_0^1 |\psi(x)|^2 dx = 1 \quad (8)$$

Note that the above equation follows from the definition. Now that we have defined what a wavelet is and a specific basis (i.e. the Haar basis), we will discuss how we can use wavelets to compute the signal to noise ratio. First, we will perform a wavelet expansion of  $f(x)$ . Let  $F(x)$  be the approximation (using wavelets) for  $f(x)$ .  $F(x)$  is defined as the following:

$$F(x) = \sum_{k,n=0}^{\infty} a_{k,n}\psi_{k,n}(x) \quad (9)$$

where  $\psi_{k,n}(x)$  is defined as in equation (6) and  $a_{k,n}$  is defined as the following:

$$a_{k,n} = \int_0^1 f(x)\psi(2^kx - n)dx \quad (10)$$

A question that arises immediately is the following: Why are we taking sum in equation (9) from  $[0, \infty)$  instead of  $(-\infty, \infty)$ ?

Recall that we had specified that  $f$  should vanish off  $[0,1]$  and that  $\int_0^1 f = 0$ . Since the aforementioned two conditions are true, all  $a_{k,n} = 0$  for negative values of  $k$  and  $n$ . Hence, we are not required to evaluate the sum in equation (9) for negative values of  $n$  and  $k$ . Thus (as mentioned in the data sets section) we have changed the data so that the average value of the stock prices is zero.

## 7.2 Wavelet Approximation (Discrete case)

In the previous section, a framework for the theory for wavelet approximations is presented. However in this paper, we will be dealing with discrete data sets (stock prices or returns). So, a framework must be developed to discretize the entire theory presented in the last section. To start, we first change the term  $f$  to  $\tilde{S}(l)$  ( $1 \leq l \leq N$ ) where  $N$  is the number of data points. We will begin by defining  $a_{k,n}$  as the following:

$$a_{k,n} = \sum_{l=0}^N \psi(2^{k-1} * l/N - n) \tilde{S}(l)/N \quad (11)$$

Notice that the definitions of  $\psi(x)$ ,  $\tilde{S}(l)$  and  $N$  are the same as defined previously. We choose a specific value for  $k$  and  $n$ . In our particular computations, we chose  $1 \leq k \leq 5$  and  $1 \leq n \leq 16$ . Now,  $F(x)$  (the wavelet approximation function to  $\tilde{S}(l)$ ) is defined as the following:

$$F(x) = \sum_{k=1}^5 2^{k-1} \left( \sum_{n=1}^{2^{k-1}} a_{k,n} \psi(2^{k-1}x - n) \right) \quad (12)$$

Notice now that  $F(x)$  is a function on  $0 \leq x \leq 1$ , while our stock prices are discrete. We sample  $F(x)$  at  $N$  points. Therefore, we compute the values for  $F(1/N)$  to  $F(1)$  with a step of  $1/N$ . Thus the function  $F(x)$ , is now quite discrete.

The array of  $F(1/N) \dots F(1)$  is called the "pure" part of the original signal. Here, part of the noise that was initially in  $\tilde{S}(l)$  has been removed. Now we wish to compute the average value of the magnitude of the "pure" signal. We define this in the following manner:



$$\bar{P} = 1/N \sum_{l=1}^N F(l/N) \quad (13)$$

So the value of  $\bar{P}$  is the magnitude of the "pure" component of the signal. The definition of the average value of the magnitude of the noise is as follows:

$$\bar{I} = 1/N \sum_{l=1}^N |\tilde{S}(l) - F(l/N)| \quad (14)$$

So the value of  $\bar{I}$  is the magnitude of the noise component of the signal. Therefore the signal to noise ratio, *snr* is defined as:

$$snr = \frac{\bar{P}}{\bar{I}} \quad (15)$$

Now we would like to briefly discuss the selection of  $k$  and  $n$ . There is no exact way to determine where a fluctuation has its "pure" component of the signal and where there is the "noise" component. For this reason, we have been placing quotation marks on both pure and noise since one can not be exactly sure where this is. Since the goal of our work is to make relative comparisons of the signal to noise ratio, we select a uniform value for the largest value of  $k$ , in otherwords  $k \leq 5$  and  $n = 2^{k-1}$  which we use throughout our analysis. Since we chose the largest value of  $k$  to be 5, this implies that  $n = 16$ . We could have chosen a higher value for the upper bound of  $k$  say  $k = 7$ , and if we used it uniformly it would have been perfectly fine. However, here we chose the upper bound to be  $k = 5$ .

### 7.3 Autocorrelation function

The autocorrelation function is a statistical measure used to determine the correlation content in any function [19]. The autocorrelation function,  $C(\tau)$  is defined as the following for a function  $x(t)$ :

$$C(\tau) = E(x(t)x(t+\tau)) \quad (16)$$

Here  $E(x(t)x(t+\tau))$  is the expectation of  $x(t)x(t+\tau)$ . So if we write this in integral form it as the following:

$$C(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t + \tau)dt \quad (17)$$

Here  $T$  represents the total time and  $\tau$  represents a shift in time. Since in most physical situations, negative time doesn't make much sense, the physical approach to the autocorrelation function is defined as the following:

$$C(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^{T/2} x(t)x(t + \tau)dt \quad (18)$$

Analogous to the case in the wavelet approximation section, we need to write down a formula for the discrete case (since our data sets are time series not continuous functions). So we approximate equation 17 using a summation. We define it as the following:

$$C(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=0}^{N/2} \tilde{S}(t)\tilde{S}(t + \tau) \quad (19)$$

where  $N, \tau \in \mathbf{Z}^+$ . Here  $N$  represents the total number of data points.

From the autocorrelation function, we wish to determine the correlation length. There are several well-known criterion for determining correlation length. The purpose of the autocorrelation function is to determine how fast the correlation falls as time shifts. In the case of a function which does not have any correlation (like white noise), the autocorrelation function for this signal behaves like that of a dirac delta function. In other words, at  $\tau = 0$ , there is a maximal value of 1 and immediately afterwards, the autocorrelation function falls to zero. One measure of correlation length is to compute the first value of  $\tau$  for which  $C(\tau) = 0$ . We use the term "first" because certain signals show periodic types of behavior in the autocorrelation function thus  $C(\tau)$  strikes zero several times. We use this measure of correlation length uniformly for all the stocks that we have studied. However, we calculate the correlation length for the "pure" signal component of each stock price fluctuation.

## 8 Results and Discussions

### 8.1 Returns results and discussion

First we will show the results for the daily returns. In fig. 3, we have plotted the returns as defined in equation (4) for Dell corp. As it is clear from the figure itself, the daily returns are random. Thus when computing the signal to noise ratio or the correlation content for the return data set, it will be clear that we will be getting random values all the time. We test this hypothesis out on all the companies in our data set and observe the same result. In fig. 4, we show the distribution of returns for Microsoft Corp. Again, one observes similar behavior to that of fig. 3. Thus we conclude that daily returns are too volatile of a quantity to look in general for trends.

We have performed a 10 day and 50 day moving average on the return data sets. An  $n$  day moving average on a data set of size  $N$  means the following:

$$\text{For all } 1 \leq i \leq (N - n) \text{ compute } S_m(i) = \frac{1}{n} \sum_{l=i}^{n+i} S(l)$$

Here  $S_m(i)$  is the new data set of returns which have smoothed with an  $n$  day moving average. The intention for performing this analysis is to attempt to "smooth" out the returns data. The larger the window size of averaging (10 and 50 in our case) the "smoother" the data set should become. We show these results in fig. 5 and in fig. 6. One can clearly observe a trend in the data sets now. However, since we have "manipulated" the data by computing essentially an average, we did not perform  $snr$  and correlation length analysis on these cases. We will mention more about studying returns in the *Future work* section.

### 8.2 Price distribution results and discussions

#### 8.2.1 Signal to noise ratio results

As we had discussed in section 7.2, we will be using the wavelet approximation to determine the signal to noise ratio for various companies in our two sectors of comparisons. In fig. 7, we show a bar graph of the different  $snr$  for the "dogs of the dow" sector. Notice that Dell Corp. has the highest value which

indicates that there are less random fluctuations in their daily closing prices than that of say American Express which showed a lower signal to noise ratio.

Next in fig. 8, we show the distribution of *snr* for the "high beta stocks" sector. There are some clearly observable trends in fig. 8. McMoran Exploration is an oil and gas company which is engaged in exploration and is thus less volatile than that of Electricidade de Portugal, an electric company of Portugal which is in bankruptcy now. This is shown in the figure by the fact that the *snr* for the Electricidade de Portugal is roughly 1/8 that of the McMoran Exploration Corp.

### 8.2.2 Autocorrelation function results

As we had discussed in section 7.3, the autocorrelation function will be used to determine the correlation length. In fig. 9, we show a sample plot for the autocorrelation function of stock from the "dogs of the dow" sector. The correlation length (as defined in section 7.3) is observed at 77. So the physical interpretation of this is that after 77 days, the correlation for the Dell stock falls to zero.

Next in fig. 10, we show a sample autocorrelation plot for a sample company in the "high beta stocks" sector. Here we plot the autocorrelation function plot for Lehman Brothers Holding Inc. We observe a correlation length of 70. So, after 70 days the correlation of the daily stock prices for this company falls to zero.

As in the case of the *snr* comparisons, we would like to plot the distributions of the correlation lengths for the "dogs of the dow" and the "high beta stocks" sectors. We plot these in figs. 11 and 12 respectively. In the case of the "dogs of the dow" we notice several large correlation length values, as in the case of the Dell, McDonald's and Home Depot. The plot in fig. 11 seems to be more or less uniformly distributed. However this is not true for fig. 12. There are some high correlation values in the case in the case of McMoran Exploration, E-Trade and Lehman Brothers Holdings while the rest are quite low.

## 9 Discussions and Conclusions

We wish to first point out that the *snr* and correlation length do not measure the same quantity. The *snr* tells how much noise there is in the signal.

The correlation length was calculated after the "noise" element of the signal was removed. This is clear by the fact that the correlation length for the "random" signal that we generated had a correlation length of 7 (shown in fig. 12). When we computed the correlation length for simply the random signal (without subtracting the noise threshold), the correlation length was 1. This is expected since all correlation should fall to zero when the time shifts.

Overall, there were some surprising results. We believed that since the "dogs of the dow" were more stable companies than the "high beta stocks" we selected, the *snr* distribution and the correlation length distribution should have been uniformly higher than in the case for the "dogs of the dow" rather than the "high beta stocks". The fact that only some companies in the "high beta stocks" were lower than the "dogs of the dow" indicates that there are other factors which one needs to take into consideration when trying to observe trends.

Likewise in the case of the correlation length, we did not observe any uniform trends. It was expected that highly volatile stocks should have lower correlation lengths (since they have more random elements in them) than that of more stable stocks such as "dogs of the dow". We did not observe this uniformly (as shown in figs. 11 and 12). However, the correlation lengths for the "dogs of the dow" were close to being uniformly distributed. Of course some stocks did have lower correlation lengths than the other, but this was not as prevalent as in the case of the "high beta stocks".

When looking at both the *snr* and correlation length, some interesting similarities and differences appear. Regarding the similarities, both measures did not show high values for General Electric company. We expected this company to not undergo any drastic changes in stock prices since it has one of the lowest volatility indexes. Another similarity is that we observed that Dell Corp. had consistent high values in the signal to noise ratios and the correlation length. With regards to the "high beta stocks", we observed very low values for the Electricidade de Portugal company. This was as expected since this stock was highly volatile and the current condition of this company is rather dismal. With the exception of the performance of E-Trade, the distributions for the *snr* and the correlation lengths are very similar in nature.

Regarding the differences, we observe differences in the high *snr* values and the high correlation length values for the "dogs of the dow". For example, we observed the highest *snr* in this sector were for the Dell, Intel and

Eastman Kodak respectively. For correlation lengths, the top three values were for McDonald's, Home Depot and Dell respectively. While it is true that *snr* and correlation length do not measure exactly the same thing, they do measure similar quantities. So one would expect a similar trend in both of these results. Now for the "high beta stocks" there is a significant difference in the correlation length of E-Trade versus its *snr*. The *snr* is quite low (comparitively speaking) for E-Trade while the correlation length is second highest in the whole sector. A possible explanation for this might be that one should change the measure of the correlation length to be the value in which  $C(\tau) = 1/e$  instead of  $C(\tau) = 0$ . The reason for this is that some stocks tend to fall very slowly after the autocorrelation function crosses  $1/2$  and in many cases the  $C(\tau)$  decays very slowly towards zero. We will mention this aspect in the *Future Work* section.

The results show that both the *snr* and the correlation length are probably dependent on factors other than the historical volatility index. The trading volume of a stock for example might have an influence on both of the aforementioned measures. We hypothesize this because stocks that are highly traded during the day may have more noisy elements in their price distribution due to the fact that there are many trades. This factor will add more noise to the data. This could very well be the case for General Electric (a steady low volatile company) whose *snr* and correlation lengths were observed to be lower than that of companies such as Dell Corp. We will discuss using trading volume as a factor in the *Future Work* section of the paper.

To conclude, we do observe both low correlation length and *snr* for the stocks which seem to have the highest volatility (such as Analog Devices Inc. and Electricidade de Portugal). In the "high beta stocks" sector, the *snr* and correlation length distributions were quite similar in nature with the exception of E-Trade company. In the case of the "dogs of the dow" we did not observe a uniform trend of highest *snr* or correlation lengths than that of the "high beta stocks". This could perhaps be explained due to the fact that we used only one year as the time frame for the entire analysis. It could very well be that companies such as McMoran Exploration which traditionally have a high volatility index had a solid performance in the year 2000. Likewise a company as stable as General Electric could have some high financial perturbations which affected the randomness in the price distribution for the year 2000. One way to test this hypothesis is to perform the same analysis for longer time periods (say 5 years for example) and then see if one can observe any clear trends between the two sectors.

## 10 Future work

During this summer the foundation for analysis of a large sample size of stocks has been built. The tools of wavelet analysis and autocorrelation function have been used to measure the information based content in the distribution of the stock prices. The goal of the future work is to expand on this knowledge and study all the major sectors in the financial market. In this paper we looked at two "extreme" ends of the spectrum of sectors: low volatile and historically stable stocks (dogs of the dow) versus historically high volatile stocks. We plan to include all the companies from all the major sectors including the following (to name a few):

- pharmaceutical
- computer - software
- computer - hardware
- biotechnology
- food - processing
- financial
- steel
- oil
- gold
- international companies in the U.S Markets

Next, we plan to break down each sector based on trading volume and volatility index. We plan on computing not only the *snr* and the correlation length, but we also plan on computing the spectral index (via power spectrum analysis). We plan on choosing a uniform low frequency region in which we will fit the  $1/f^\alpha$  power law ( $\alpha$  is the spectral index). Then we plan on computing the correlation coefficient between the trading volume and the following measures: the *snr*, correlation length and the spectral index for all the stocks in the particular sector. This will help us determine how much of an influence the trading volume has on the above three measures

in the particular sector. Theoretically speaking, the influence of the trading volume could vary from sector to sector and this is why we will carry it out on all the sectors studied. Overall, we hope that we will be able to use the aforementioned three methodologies to find the similarities and or differences between the sectors.

Second, we plan on observing how the *snr* varies over time. Throughout this paper, we used a uniform time frame (1 year exactly). We plan on using several short time intervals (such as 3 months) and compute the *snr* for each interval and then plot how the *snr* varies as a function of time. We plan on determining this for select number of companies from each sector. We hope to see if we can observe any trends of the *snr* varying over time.

Third, we plan on computing the *snr*, correlation length and the spectral index for two different time periods. In this paper we computed the mathematical measures for a one year time period. We plan on computing the above measures for a one year time period and for a longer time period (say 5 years). We believe that we should be able to observe some distinct differences between the two extreme sectors (dogs of the dow and high beta stocks) at least. Perhaps, interesting trends will be observed between other sectors (not just the extreme ends).

Finally, we plan on looking at the distribution of daily returns after 50 day averaging for the stocks among the two sectors studied in this paper. Since performing a 50 day average inherently changes our data set, we will carefully examine particular cases of *snr*, correlation length and spectral index for the daily returns. If one can observe clear distinct trends among the two extreme sectors, we will plan on pursuing this area as we did with the stock price fluctuations.

## 11 References

- [1] S.V. Buldyrev, N.V. Dokholyan, A.L. Goldberger, S. Havlin, C.K. Peng, H.E. Stanley, G.M. Viswanathan, *Physica A* **249** 1998 430.
- [2] S. Guharay, B.R. Hunt, J.A. Yorke, O.R. White, *Physica D* **146** 2000 388.
- [3] J. Barrel P., A. Hasmy, J. Jiménez, A. Marcano, *Physical Review E* **61** 2000 1812.
- [4] P. Bernaola-Galván, I. Grosse, P. Carpena, J.L. Oliver, R. Román-Roldán, H.E. Stanley, *Physical Review Letters* **85** 2000 1342.
- [5] W. Li, T. Marr, K. Kaneko, *Physica D* **75** 1995 217.



- [6] F. Black, M. Scholes, *Journal of Political Economy* **81** 1973 637.
- [7] J-F Muzy, D. Sornette, J. Delour, A. Arneodo, *Quantitative Finance* **1** 2001 131.
- [8] A. Arneodo, J-F Muzy, D. Sornette, *Eur. Phys. J. B* **2** 1998c 277.
- [9] J-F Muzy, J. Delour, E. Bacry, *Eur. Phys. J. B* **17** 2000 537.
- [10] J. Laherrère, D. Sornette, *Eur. Phys. J. B* **2** 1998 525.
- [11] V. Plerou, P. Gopikrishnan, L.A. Nunes Amaral, M. Meyer, H.E. Stanley, *Physical Review E* **60** 1999 6519.
- [12] B.B. Mandelbrot, *Sci. Am.* **280** February 1999 70.
- [13] B.B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk: Selecta vol. E*, Springer-Verlag, New York, 1997.
- [14] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge, 2000.
- [15] N. Vandewalle, M. Ausloos *Eur. Phys. J. B* **4** 1998 257.
- [16] Details can be found at <http://wrds.wharton.upenn.edu>
- [17] R. Cont, *Scaling and correlation in financial data*, Preprint cond-mat/9705075, 1999.
- [18] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
- [19] D.E. Newland, *An Introduction to Random Vibrations, Spectral and Wavelet Analysis*, Longman Scientific & Technical, New York, 1993.

## 12 Acknowledgements

I would like to thank my advisor, Prof. Joseph Stampfli for his continuous guidance and suggestions. Also thanks are due to Mr. Craig Eichen for his helpful suggestions for obtaining the data sets. The other students of the REU this year, namely, Jiho, Ruth, Erika, Danielle, Julia, Elizabeth, Jonathan C., Jonathan L., John and Dan deserve thanks for constructive discussions and other forms of academic assistance. Finally, last but not at all the least, invaluable support and encouragement from my parents has been a guiding light throughout my research work.

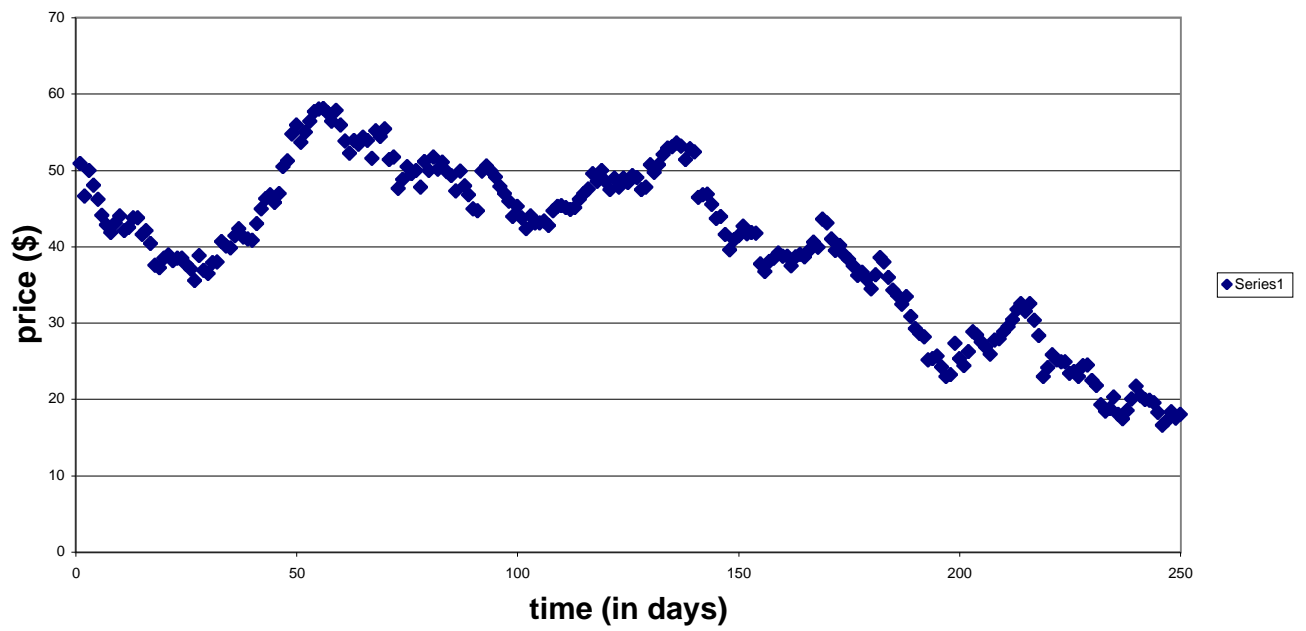


Figure 1: Distribution of daily closing prices for Dell Corp. from Jan. 00- Dec.00

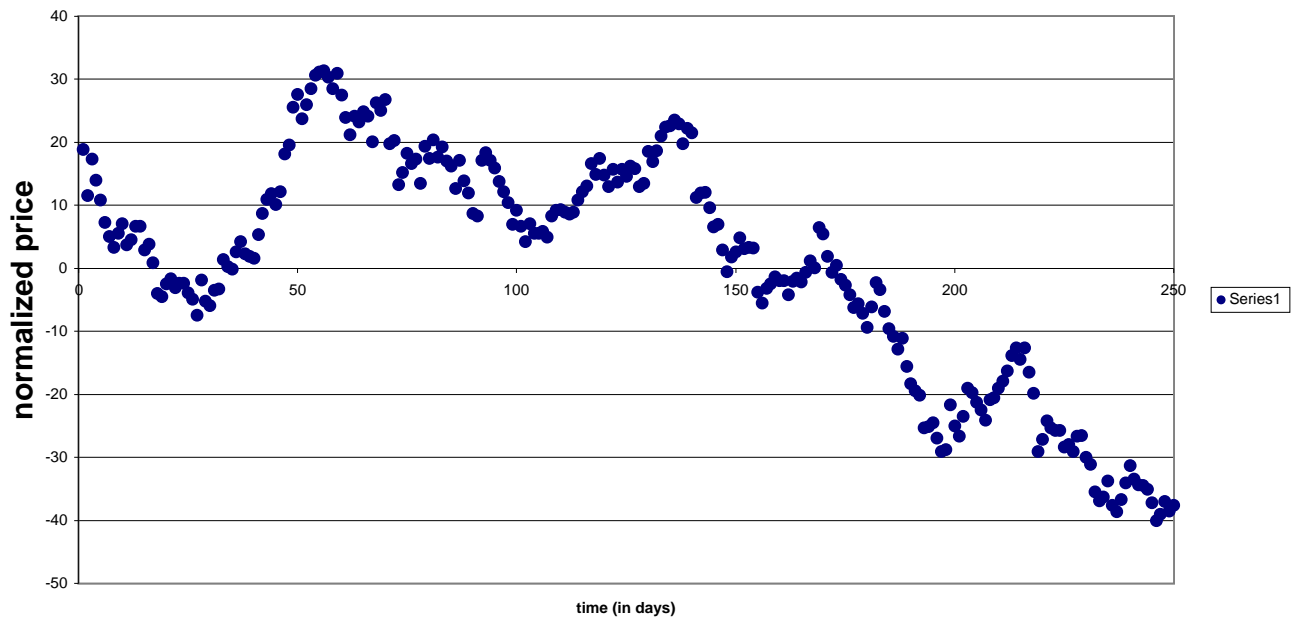


Figure 2: Distribution of normalized prices for Dell Corp. from Jan. 00 - Dec. 00

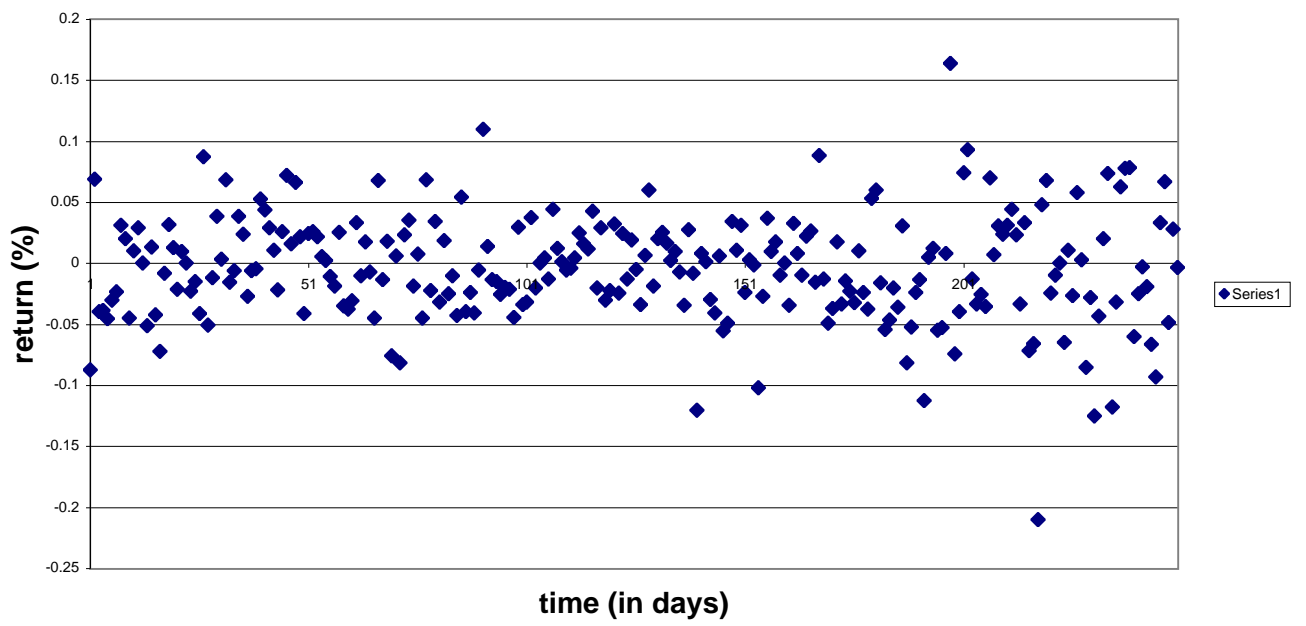


Figure 3: Distribution of daily returns for Dell Corp. from Jan. 00-Dec.00

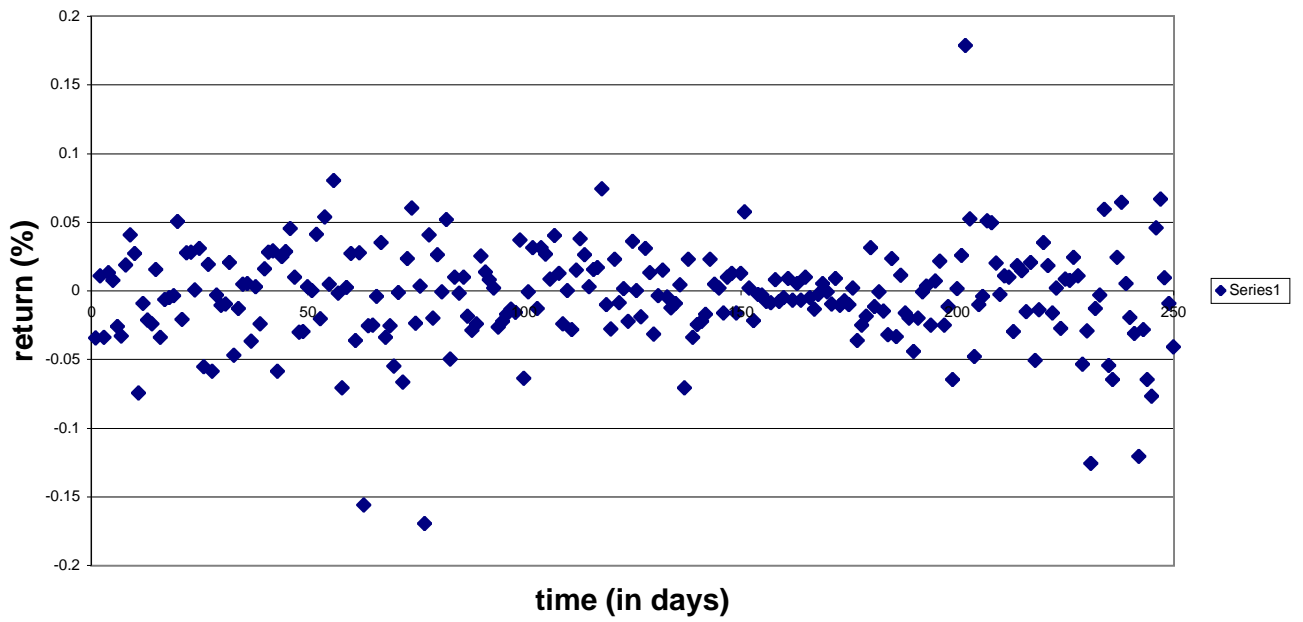


Figure 4: Distribution of daily returns for Microsoft Corp. from Jan. 00 - Dec. 00

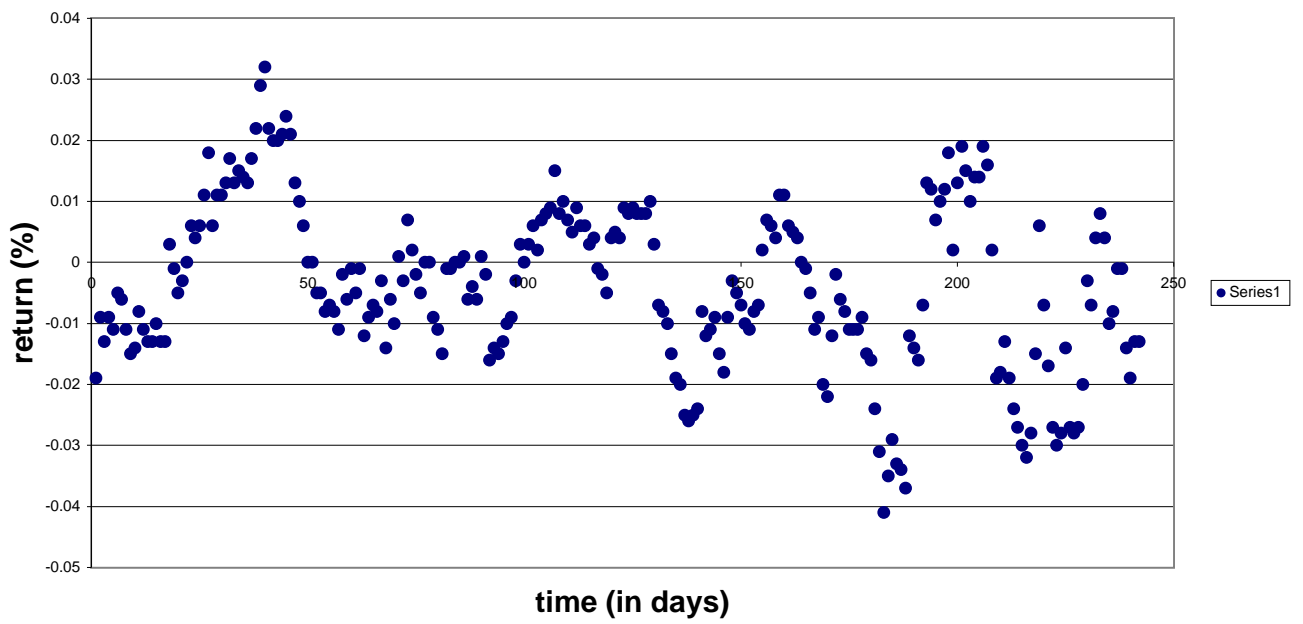


Figure 5: Distribution of daily returns for Dell Corp. (after applying 10 day moving average)

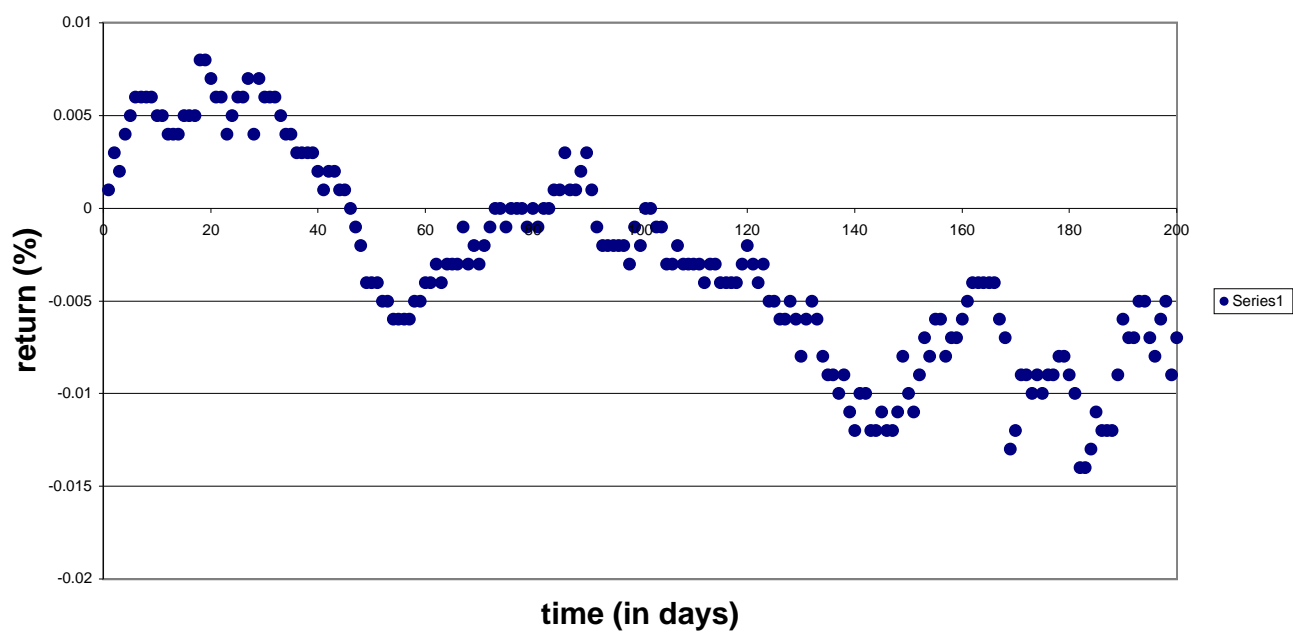


Figure 6: Distribution of daily returns for Dell Corp. (after applying 50 day moving average)

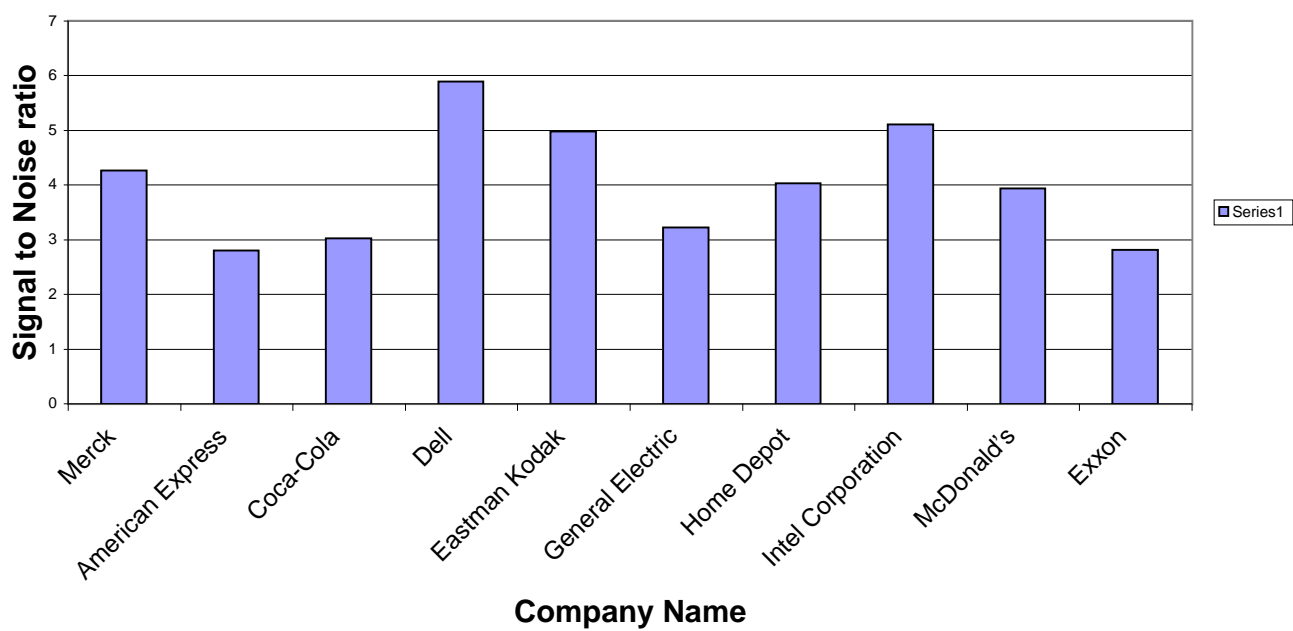


Figure 7: Distribution of signal/noise for "Dogs" of the Dow



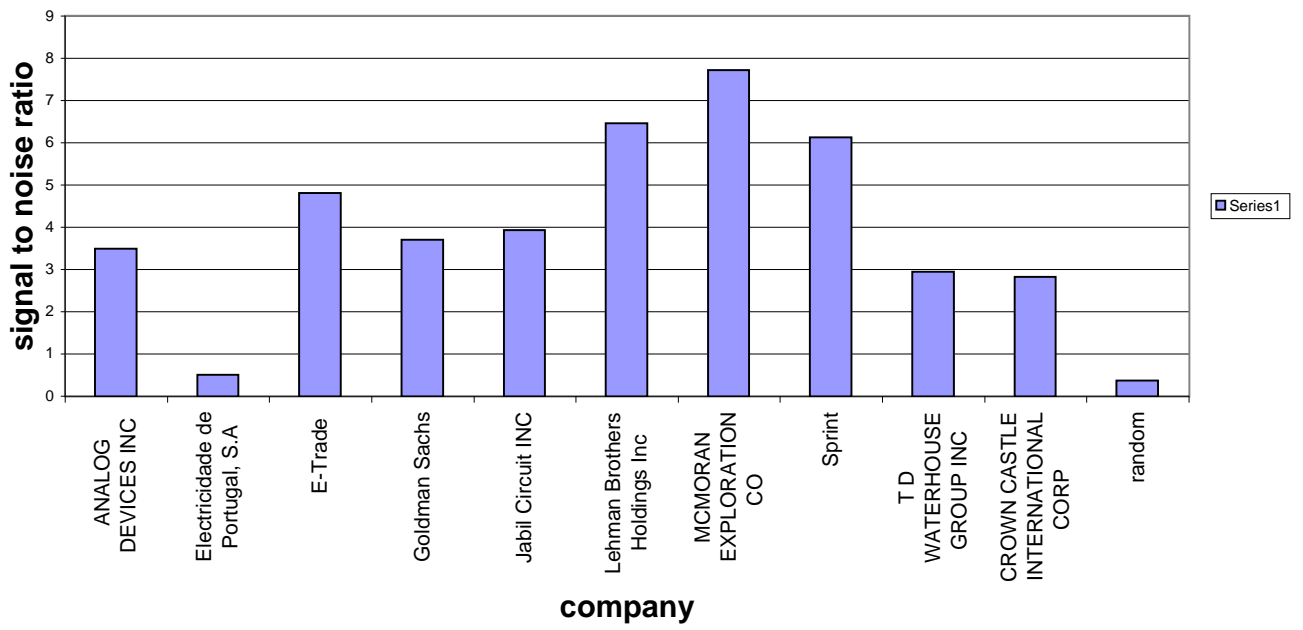


Figure 8: Distribution of signal/noise for high volatile stocks

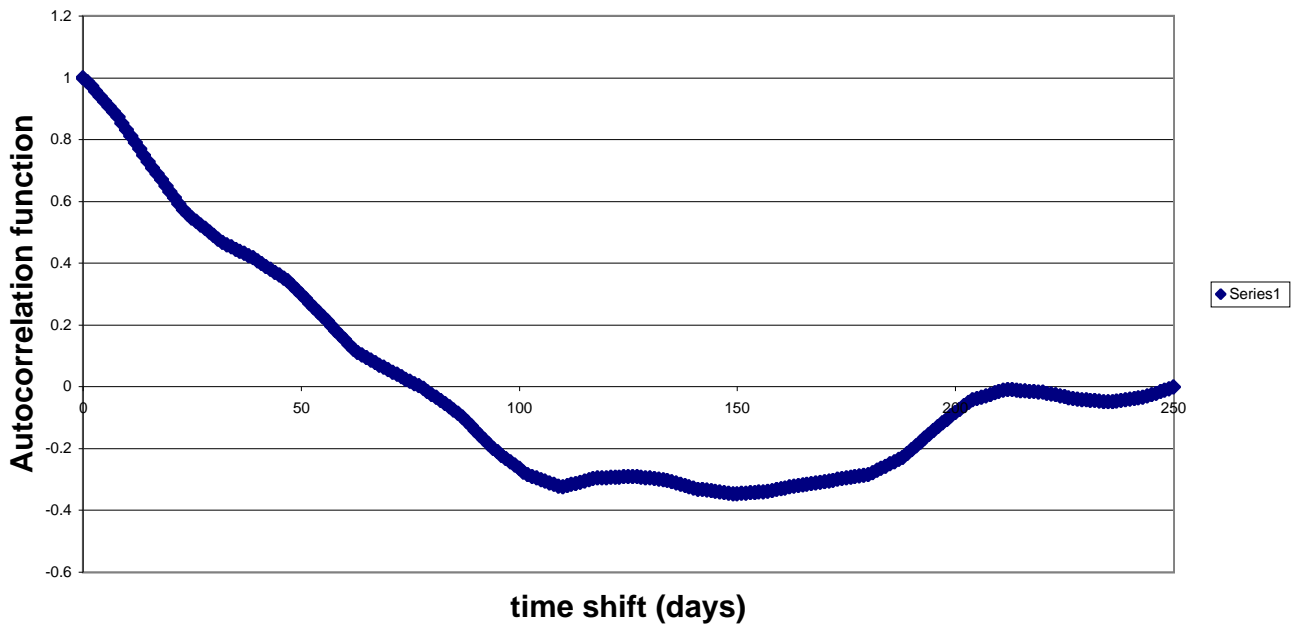


Figure 9: Autocorrelation function plot for Dell Corp.)



Figure 10: Autocorrelation function plot for Lehman Brothers Corp.)

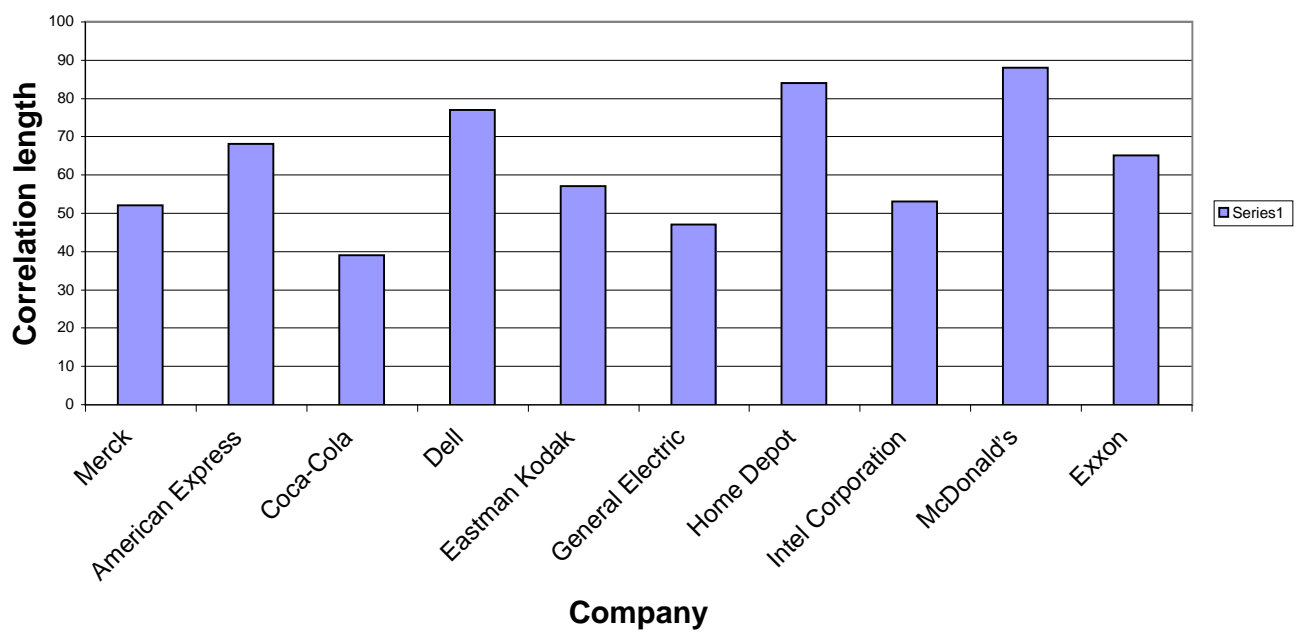


Figure 11: Distribution of correlation lengths for "dogs" of the dow

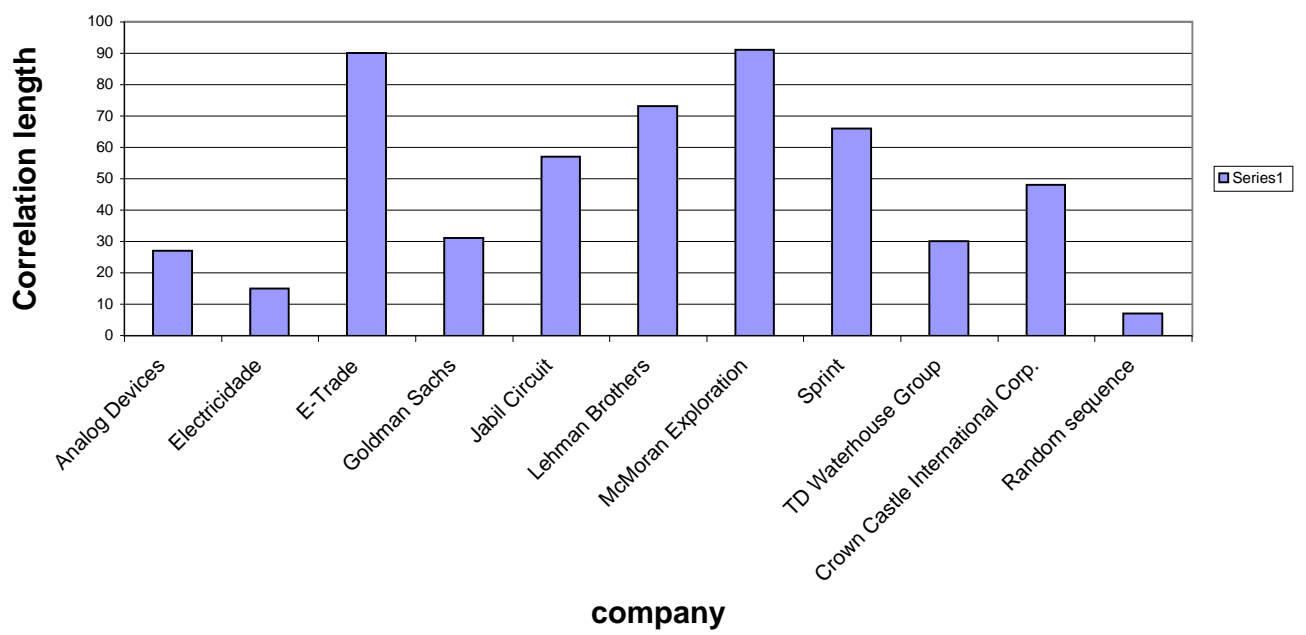


Figure 12: Distribution of correlation lengths for highly volatile stocks