Chapter 3

3.1. Introduction

Statistical and Fourier analysis methods for the time series representing physical phenomena from various fields such as biology, geophysics, meteorology, finance mathematics, DNA sequence analysis, seismic data, rainfall data, temperature variation data, ECG data and price fluctuations of different stocks in different time intervals have been quite popular in the scientific world. With the surge of data outpouring from the various fields, it is becoming imperative to develop and use proper mathematical tools for classification and understanding those systems. These developments have motivated researchers to apply some classical methodologies to understand dynamics of stock market fluctuations. A tremendous amount of work using wavelet applications (Muzy et al. (2000, 2001), Arneodo et al. (1999c) and L’evy distributions and spectral analysis (Arneodo et al. (1992) and Arneodo et al. (2002)), multifractal models (Mandelbrot (1997, 1999) and Mantegna and Stanley (2000)) (to name a few) have been used as tools to study the financial markets.

In general, most of this research work has been applied to measure the fluctuations of market indices. For example, a lot of work has been done in efforts to detect the trends in the S&P 500 (for various different time periods). The work has been done on currency exchange dynamics also (Vandewalle and Ausloos (1998)). In most of these cases correlation content has been measured for long term periods (lengths of at least five years). There seems to be dearth of comprehensive work on looking at differences in actual stock prices. More recently, analysts in the Wall Street have started using some of the mathematical methods for their own analysis. A natural curiosity is whether these methods could be used to analyze and understand fluctuations of Indian and Saudi stocks. We know that the stock market analysts in Mumbai are using some software to keep financial data record and analyze it. However wavelet and fractal methods are not used.

The distribution of price fluctuations is important from theoretical point of view, and helpful in understanding dynamics of the stock market. In this chapter, we focus our study on distribution of daily closing prices, distribution of normalized
prices, distribution of daily returns, signal to noise ratio, autocorrelation function, and autocorrelation length for BSE 100, Nifty 50, stock of Riyadh Bank and stock of British Bank.

BSE 100 and Nifty data sets discussed in this chapter have been taken from Viratech software solution Pvt. Ltd. Saudi stock data set was taken from Siddiqi and Ahmad (2001). We consider the daily closing prices for the time period April 2003 – March 31, 2004 (The days on which the stock exchanges were open). This generated a data set of 256 points. We study the Riyadh Bank and British Bank stocks for the period April-May 2000. MATLAB wavelet toolbox and a programme written by Mr. S.D. Khan of KFUPM (King Fahd University of Petroleum and Minerals) have been used to study the BSE 100, Nifty 50, Stocks of British Bank (BB) and Riyadh Bank (RB) for the period mentioned above. The work in this chapter is motivated by Guharay (2002) where S & P 500 stocks have been analyzed by wavelet tools.

This chapter describes the work done on use of mathematical methods for modeling price fluctuations of financial time series. This chapter has been divided into six sections. The first section is an introductory section. In the second section we explain some basic terms like stock price distribution and daily returns of the stock market. We have divided the third section into two subsections in which we discuss wavelet approximation to a continuous function, the discrete case and the signal to noise ratio of a signal. In the fourth section, we analyze the time series data with MATLAB wavelet toolbox where as in the subsection, we have discussed autocorrelation function and correlation length of BSE 100, Nifty 50, British Bank and Riyadh Bank. Results and discussion are given in the fifth section. In the last section we give conclusions to the analysis performed. Results of this chapter are published in Manchanda, P. et al. (2007b).

3.2. Preliminaries

Definition 3.2.1. For splitting of stocks, we define \( f_{acpr} \) as the number of additional shares per old share issued

\[
f_{acpr} = \frac{s(t) - s(t')} {s(t')},
\]
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 the date before the split.

We wish to analyze two separate events:

(i) The distribution of prices.

(ii) The distribution of daily returns.

**Definition 3.2.2** Classically, we define the *stock price distribution* as

\[
S(t) = S_0 e^{\mu t + \sigma W_t},
\]

(3.2)

where

- \(S_0\) is the initial price of the stock,
- \(\mu\) is a measure of the average rate of growth of the stock price, also called drift,
- \(\sigma\) is the volatility which measures the standard deviation of the returns,
- \(t\) is time, and
- \(W_t\) represents the Brownian motion drift term which is normally distributed with mean 0 and standard deviation of \(\sqrt{t}\).

Defining the daily return \(r\) as

\[
\frac{S(t + \Delta t)}{S(t)}
\]

(3.3)

(where \(\Delta t = 1\) day) it is clear that after taking logarithms in equation (3.2) we get

\[
log \frac{S(t + \Delta t)}{S(t)} \propto \mu \Delta t.
\]

(3.4)

For the sake of computational simplicity, we define the *daily return* as

\[
r = log \frac{S(t + \Delta t)}{S(t)}.
\]

(3.5)

We obtain the entire daily closing stock prices for each of the four mentioned stocks. This was done for a period of one year, 1st April 2003 to 31st March 2004, for BSE 100 and Nifty 50 and for the period, April-May 2000 for Riyadh Bank (RB) and British Bank (BB) stocks.

These closings can be normalized using (3.1). Let us call these closings as \(S(t)\) \((1 \leq t \leq 256)\). However very often the following normalization is adapted.

For all values of \(t\) for a particular company, we first determine

\[
\alpha = \max_t S(t).
\]

(3.6)
Next, for all values of $t$, we define,

$$\tilde{S}(t) = \frac{100S(t)}{\alpha}. \tag{3.7}$$

The above steps ensure that all stock closings are distributed between $(0, 100)$. We perform one more scaling on $S(t)$. For simplicity in the same future calculations, we require the following final condition on stock prices:

$$\sum_{t=1}^{n} S(t) = 0. \tag{3.8}$$

Thus we subtract the average value of all prices from each price to ensure that the equation (3.8) holds. We call these final prices as *normalized prices* denoted by $\tilde{S}(t)$.

![Figure 1: BSE 100 closing prices.](image)
Figure 2: BSE 100 normalized data.

Figure 3: Daily closing of Nifty 50.
Figures 1-4 show distribution of daily closing and normalized closing of BSE 100 and Nifty 50. Figures 5-8 show the distribution of closing prices and normalized prices for the stocks of Riyadh Bank (RB) British Bank (BB) for the said period of time. Here the plots of $\tilde{s}(t)$ have been drawn against $t$. There may be negative values of $\tilde{s}(t)$, however the trend of fluctuation is preserved. It may be remarked that for smaller size of data normalization and scaling is not necessary. For a statistical analysis of Saudi stock markets we refer to Siddiqi and Ahmad (2001) and for a simple introduction of the role of Mathematics in financial problems we cite Wilmot (1998).
Figure 5: Daily closing of RB.

Figure 6: RB normalized data.
Figure 7: Daily closing of BB.

Figure 8: BB normalized data.
Figure 9: BSE 100 returns in volatile form.

Figure 10: Nifty 50 returns in volatile form.
Figure 11: RB returns in volatile form.

Figure 12: BB returns in volatile form.
3.3. **Mathematical Methods**

We apply two mathematical methods, namely, Haar wavelet approximation and autocorrelation function analysis on the daily returns of the stock data. Using the wavelet approximation, we compute the signal to noise ratio \((\text{snr})\). We use the autocorrelation function to compute the correlation length.

![BSE 100 returns graph.](image1)

![Nifty 50 returns graph.](image2)
3.3.1. Wavelet Approximation to Continuous Function
Let $f$ be a function (signal) with finite energy, that is, $\int_{-\infty}^{\infty} |f(t)|^2 dt < \infty$.

In mathematical terminology $f(t) \in L_2(R)$.

Suppose $\psi$ is a real valued function having compact support such that

$$\int_{-\infty}^{\infty} \psi(t) dt = 0.$$ 

We define

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j, k \in \mathbb{Z} \quad (3.9)$$

Then $\psi$ is called a wavelet (more precisely mother wavelet) if $\{\psi_{j,k}(t)\}$ is an orthonormal basis for $L_2(R)$.

One can use $\langle f, \psi_{j,k} \rangle$ ($\langle \cdot, \cdot \rangle$) denotes the inner product on $L_2(R)$ which is defined as $

\langle f, g \rangle = \int f(t) g(t) dt$

to determine the original function.

The simplest example of the wavelet is the Haar function which is Daubechies-1 wavelet, see, for example, Daubechies (1992) and Siddiqi (2004).

We recall here that the Haar wavelet is defined by the following:

$$\psi(t) = \begin{cases} 1; & 0 \leq t \leq 1/2, \\ -1; & 1/2 \leq t \leq 1, \\ 0; & \text{otherwise}. \end{cases}$$

It is clear that

$$\int_{0}^{1} |\psi(t)|^2 dt = 1 \quad (3.10)$$

Now we indicate how wavelet is used to compute the signal to noise ratio (SNR).

For this, let $F(t)$ be the approximation (using wavelet) for $f(t)$, that is,

$$F(t) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} a_{j,k} \psi_{j,k}(t), \quad (3.11)$$

where $\psi_{j,k}(t)$ is defined as in equation (3.9) and $a_{j,k}$ is defined as the following:

$$a_{j,k} = \int_{0}^{1} f(t) \psi(2^j t - k) dt. \quad (3.12)$$

For Haar wavelet we have $\psi$ equal to zero outside the $[0, 1]$ and $\int_{-\infty}^{\infty} \psi(t) dt = 0$, these two conditions imply that $a_{j,k} = 0$ for negative values of $j$ and $k$, thus we have taken (3.11) from 0 to $\infty$ instead of $(-\infty, \infty)$.

3.3.2. Wavelet Approximation (Discrete Case)
In this section we discuss wavelet approximation of discrete sets (stock prices or daily returns). We first convert the continuous function to discrete function, say \( \tilde{S}(l) \), \( 1 \leq l \leq n \), where \( n \) is the number of data points. We define \( a_{j,k} \) as

\[
a_{j,k} = \sum_{i=1}^{n} \psi \left( 2^{j-1} \frac{l}{n} - k \right) \tilde{S}(l)/n. \tag{3.13}
\]

Here \( \psi(x) \), \( \tilde{S}(l) \) and \( n \) are as mentioned earlier.

Now \( F(t) \), the wavelet approximation to \( \tilde{S}(l) \) is defined as the following. For \( 1 \leq j \leq 5 \), \( 1 \leq k \leq 16 \):

\[
F(t) = \sum_{j=1}^{5} 2^{j-1} \left( \sum_{k=1}^{2^{j-1}} a_{j,k} \psi(2^{j-1}t - k) \right). \tag{3.14}
\]

We notice that \( F(t) \) is a function on \([0, 1]\) while our stock prices are discrete. We sample \( F(t) \) at \( n \) points. Therefore, we compute the values for \( F \left( \frac{1}{n} \right) \) to \( F(1) \) so that \( F(t) \) is now discrete. The array \( F \left( \frac{1}{n} \right), F \left( \frac{1}{n-1} \right), ..., F(1) \) is called the pure part of the original signal. Here the noise part that was initially present in \( \tilde{S}(l) \) has been removed.

The average value of the magnitude of the pure signal is defined as follows:

\[
\bar{p} = \frac{1}{n} \sum_{l=1}^{n} F \left( \frac{1}{l} \right). \tag{3.15}
\]

This is called the magnitude of the pure component. The average value of the magnitude of the noise is defined to be

\[
\bar{q} = \frac{1}{n} \sum_{l=1}^{n} \left| \tilde{S}(l) - F \left( \frac{l}{n} \right) \right|, \tag{3.16}
\]

\( \bar{q} \) is called the magnitude of the noise component of the signal.

The signal to noise ratio abbreviated as \( snr \) is defined as

\[
\text{snr} = \frac{\bar{p}}{\bar{q}}. \tag{3.17}
\]

A time series or signal representing the graph of \( t \) versus \( S(t) \) has two components, one is called ‘pure component’ and the other is called ‘noise component’ (unwanted contents of the signal). The \( \text{snr} \) tells how much noise is there in the signal. It may be used as a characteristic of the signal. Two signals may be compared through their \( \text{snr} \)
values. The value of $snr$ of a signal is an indicator of fluctuations in its daily closing prices. Smaller $snr$ indicates less random fluctuations.

### 3.4. MATLAB Toolbox Wavelet Analysis

The discrete wavelet analysis of the time series data under consideration is performed in terms of decomposition, approximation and denoising of the original signals. The decomposition and statistical analysis of financial time series data for BSE 100, Nifty 50, British Bank and Riyadh Bank is shown in figures 17 to 24. Fig.17 shows the time series analysis of BSE 100 with Haar wavelet at level 3. This figure has five parts. The first part “S” represents the time series signal or raw data. The second part $a_3$ corresponds to the amplitude of the signal for the Haar wavelet at level 3. The last three parts $d_1$, $d_2$ and $d_3$ of the figure represent detail of the signal at three different levels.

![Wavelet Analysis Diagram](image)

Figure 17: BSE 100 data analysis using MATLAB Haar wavelet, level 3.
Figure 18: BSE 100 statistical analysis using MATLAB - Haar wavelet, level 3.

Figure 19: BSE 100 – Original /de-noised data: using MATLAB - Haar wavelet, level 3.
Similarly, in figure 20, time series data of the Nifty 50 has been analyzed by using db4 wavelet at level 4. At this level we get six parts in the figure with first part “S”
representing the time series signal of Nifty 50, the second part $a_4$ giving the signal amplitude and rest $d_1$ to $d_4$ showing signal details at four different levels.

A close look at the signal $S$ and amplitude parts in figures 17 and 20 and plots of their pure parts in figures 25-26 indicates that BSE 100 and Nifty 50 show almost similar closing patterns.

Figure 22: Nifty 50–Original /de-noised signal: Daubechies 4 wavelet at level 4.

Figure 23: RB statistical analysis (April/May 2000) using Haar wavelet level 2.
Figure 24: BB statistical analysis (April/May 2000) using MATLAB-Haar wavelet level 2.

Figures 19 and 22 exhibit the wavelet denoising of BSE 100 and Nifty 50. Each figure contains original details coefficients, original and denoised signals, original coefficients and threshold coefficients. For a glimpse at the statistical properties of the data under consideration, we used Haar wavelet at level 3 for BSE 100 (Fig.18) and db4 wavelet at level 4 for Nifty 50 (Fig.21). These statistical details includes histogram, cumulative histogram, autocorrelation and fast Fourier transform spectrum. The statistical analysis of Bank of Riyadh and British Bank have been shown in Fig. 23 and Fig. 24 respectively. Haar wavelet at level 2 has been used for this analysis. Figures 25-28 shows pure part of the time series signals of BSE 100, Nifty 50, Bank of Riyadh and British Bank.
Figure 25: BSE 100 – pure part.

Figure 26: Nifty 50 – pure part.
Figure 27: RB – pure part.

Figure 28: BB – pure part.
3.4.1. Autocorrelation Function

We recall that a data set exhibits *auto correlation* if the value \( x_i \) at time \( t_i \) is correlated with the value \( x_{i+d} \) at time \( t_{i+d} \) where \( d \) is some time increment in the future. In a long memory process, autocorrelation decays over time and this decay follows a power law namely

\[
p(k) = Ck^{-\alpha},
\]

where \( C \) is a constant and \( p(k) \) is the autocorrelation function with lag \( k \).

For the given data \( X_1, X_2, X_3, ..., X_n \) at time \( t_1, t_2, t_3, ..., t_n \), we recall that the lag \( k \) autocorrelation function is

\[
r_k = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2},
\]

where

\[
\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}.
\]

Figure 29: BSE 100 – autocorrelation.
The data observations used here are equally spaced. Autocorrelation is a correlation coefficient. However, instead of correlation between two different variables, autocorrelation is the correlation between two values of the same variable at time $t_i$ and $t_{i+d}$.

The autocorrelation function is being used to detect the non-randomness in the data and to identify an appropriate time series model if data is not random. When autocorrelation is used to detect non-randomness, it is only usually the first (lag 1) autocorrelation that is of interest. When the autocorrelation is used to identify appropriate time series model, the autocorrelations are plotted for many lags.

![Figure 30: Nifty 50 – autocorrelation.](image)

In general the autocorrelation function $C(h)$ (Mantegna and Stanley (2000)) is defined as

$$C(h) = \lim_{T \to \infty} \frac{1}{T} \int_0^{T/2} x(t)x(t+h)dt.$$  \hspace{1cm} (3.21)

In the discrete case it is defined as

$$C(h) = \lim_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n/2} \tilde{x}(t)\tilde{x}(t+h),$$  \hspace{1cm} (3.22)
where $h, n$ are positive numbers. Here $n$ represents the total number of data points. From autocorrelation function, we wish to determine correlation length.

![Figure 31: RB – autocorrelation.](image)

We use autocorrelation function to determine how fast the correlation falls as the time shifts. In case of a function (signal) which does not have any correlation (like white noise), the autocorrelation function for this signal behaves like Dirac delta function. In other words, at $h = 0$, there is a maximal value of 1 and immediately afterwards, the autocorrelation falls to zero. One measure of correlation length is to compute the first value $h$ for which $C(h) = 0$. We use the term first because certain signals show periodic behavior in the autocorrelation function, thus $C(h)$ may take several zero values. We used this measure of correlation length for all data under consideration.

We calculate the correlation length for pure signal component of each stock price fluctuation.

The cross-correlation coefficient $r$ which is a measure of linear association between two variables is defined as

\[
r = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}.
\]  

(3.23)
A positive value of coefficient $r$ indicates that as one value increases the other tends to increase, whereas a negative value of $r$ indicates that as one variable increases the other tends to decrease.

### 3.5. Results and Discussion

First we plot the results for the daily returns for BSE 100, Nifty 50 for one year and for stocks of Riyadh Bank and British Bank for two months, see figures 9-12. As clear from figures 9-12, the daily returns are random. So we expect that the signal to noise ratio ($snr$) or correlation content for the return data set will give random values. It is clear that the daily returns are too volatile a quantity to look in general for trends.

#### 3.5.1. Experimental Results

MATLAB wavelet toolbox analysis of BSE 100 data by Haar wavelet at level 3 is given in Figure 17. Figure 20 shows MATLAB wavelet toolbox analysis of Nifty 50 data by use of Daubechies 4 wavelet at level 4. The experimental results are given in figures 18-24.
3.5.2. Signal to Noise Ratio Results

It may be remarked that $\bar{P}$, $\bar{I}$ and $snr$ have been computed for normalized data of stocks under consideration. Figures 25-28 provide pure parts of the signal of data, respectively, BSE 100, Nifty 50, Riyadh Bank and British Bank.

3.5.3. Autocorrelation Function Results

The autocorrelation function is used to determine the correlation length. Figures 29-32 show plot for the autocorrelation function of stocks BSE 100, Nifty 50, Riyadh Bank and British Bank discussed above.

In Fig. 29, we observed a correlation length of 82 for BSE 100 which shows that after 82 days the correlation of the daily closings for this market falls to zero. Next in Fig. 30, we show autocorrelation plot for Nifty 50. Here we observe a correlation length of 80. So after 80 days the correlation of daily closings of NSE falls to zero.

The correlation length of four stocks is given in Table 1

**Table 1:** $\bar{P}$, $\bar{I}$ and $snr = \frac{\bar{P}}{\bar{I}}$ for BSE 100, Nifty 50, BB and RB.

<table>
<thead>
<tr>
<th>Stocks</th>
<th>$\bar{P}$</th>
<th>$\bar{I}$</th>
<th>$snr$</th>
<th>C.L</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSE 100</td>
<td>-29.5275</td>
<td>4.4885</td>
<td>-6.5785</td>
<td>82</td>
</tr>
<tr>
<td>Nifty 50</td>
<td>-31.1585</td>
<td>4.4781</td>
<td>-6.9580</td>
<td>80</td>
</tr>
<tr>
<td>BB</td>
<td>-9.3043</td>
<td>0.0753</td>
<td>-123.5630</td>
<td>9</td>
</tr>
<tr>
<td>RB</td>
<td>-7.4354</td>
<td>0.0803</td>
<td>-92.5952</td>
<td>10</td>
</tr>
</tbody>
</table>

C.L = Correlation length

3.6. Conclusion

We wish to first point out that the signal to noise ratio and correlation length of a signal is indicator of its fluctuations. In the case of correlation length it is expected that high volatile stocks should have lower correlation lengths (since they have more random elements in them) than that more stable stocks. However, it has not been observed in the case of S & P 500 (Guharay, S. (2002)) and similar is the case in our study, see Table 1. BSE 100 is more volatile than the British Bank (BB) stock but the
correlation length of BSE 100 is 82. In view of this it is essential to investigate further the factors influencing the correlation length. The correlation length was calculated after the ‘noise’ element of the signal was removed. It has been observed that BSE 100 and Nifty 50 have similar pattern of fluctuations. Their $snr$ calculated here are approximately the same for the period under consideration which confirms this hypothesis and consequently correctness of our procedure is apparent.

***************