Evaluating the Predictability of Financial Time Series A Case Study on Sensex Data

P. Bagavathi Sivakumar [1], Dr. V. P. Mohandas [2]
[1] Assistant Professor, Department of Computer Science and Engineering,
[2] Professor and Chairman, Department of Electronics and Communication Engineering,
School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore – 641 105, India.

Abstract:

A discrete- time signal or time series is set of observations taken sequentially in time, space, or some other independent variable. Examples occur in various areas including engineering, natural sciences, economics, social sciences and medicine. Financial time series in particular are very difficult to model and predict, because of their inherent nature. Hence, it becomes essential to study the properties of signal and to develop quantitative techniques. The key characteristics of a time series are that the observations are ordered in time and that adjacent observations are related or dependent. In this paper a case study has been performed on the BSE and NSE index data and methods to classify the signals as Deterministic, Random or Stochastic and White Noise are explored. This pre-analysis of the signal forms the basis for further modeling and prediction of the time series.

Keywords:
Time Series, Signal Analysis, Time Series Analysis, Deterministic, Stochastic.

I. Introduction

A discrete-time signal or time series [1] is a set of observations taken sequentially in time, space or some other independent variable. Many sets of data appear as time series: a monthly sequence of the quantity of goods shipped from a factory, a weekly series of the number of road accidents, hourly observations made on the yield of a chemical process and so on. Examples of time series abound in such fields as economics, business, engineering, natural sciences, medicine and social sciences.

An intrinsic feature of a time series is that, typically, adjacent observations are related or dependent. The nature of this dependence among observations of a time series is of considerable practical interest. Time Series Analysis is concerned with techniques for the analysis of this dependence [2]. This requires the development of models for time series data and the use of such models in important areas of application.

A discrete-time signal x (n) is basically a sequence of real or complex numbers called samples. Discrete-time signals can arise in various ways. Very often, a discrete-time signal is obtained by periodically sampling a continuous-time signal, that is $x(n) = x_c(nT)$, where $T = 1 / F_s$ is the sampling period and F_s is the sampling frequency. At other times, the samples of a discrete-time signal are obtained by accumulating some quantity over equal intervals of time, for example, the number of cars per day traveling on a certain road. Financial signals, like daily stock market prices are inherently discrete-time.

When successive observations of the series are dependent, the past observations may be used to predict future values. If the prediction is exact, the series is said to be deterministic. We cannot predict a time series exactly in most practical situations. Such time series are called random or stochastic, and the degree of their predictability is determined by the dependence between consecutive observations. The ultimate case of randomness occurs when every sample of a random signal is independent of all other samples. Such a signal, which is completely unpredictable, is known as White noise and is used as a building block to simulate random signals with different types of dependence. To properly model and predict a time series, it becomes important to fundamentally and thoroughly analyze the signal it self, and hence there is a strong need for signal analysis.

II. Signal Analysis

The classification of signals as deterministic, random or stochastic and white noise is very important in deciding about models and methods for prediction. The signal analysis has to be viewed in this regard. The primary of goal of signal analysis is to extract useful information that can be used to understand the signal generation process or extract features that can be used for signal classification purposes. Typical applications of signal analysis include detection and classification of radar and sonar targets, speech and speaker recognition, detection and classification of natural and artificial seismic events, event detection and classification in biological and financial signals, efficient signal representation for data compression, image processing, etc.

A. Signal analysis Techniques

The main objective of signal analysis is the development of quantitative techniques to study the properties of a signal and the differences and similarities between two or more signals from the same or different sources. The major areas of random signal analysis are:

1. Statistical analysis of signal amplitude, that is the sample values

T. Sobh (ed.), Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, 99–104.

2. Analysis and modeling of the correlation among the samples of an individual and
3. Joint signal analysis that is, simultaneous analysis of two signals in order to investigate their interaction or interrelationships.

The various random signal analysis techniques found in the literature are shown in the Figure 1.

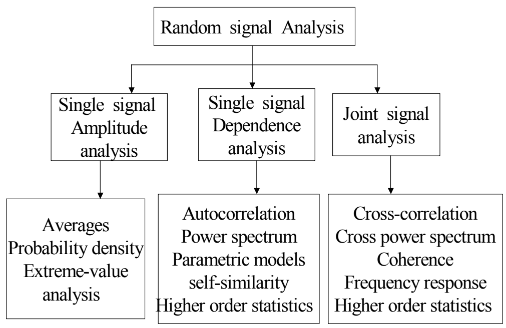

Fig. 1. **Signal analysis Techniques.**

The prominent tool in signal analysis is spectral estimation, which is a generic term for a multitude of techniques used to estimate the distribution of energy or power of a signal from a set of observations. Spectral estimation finds many applications in areas such as medical diagnosis, speech analysis, seismology and geophysics, nondestructive fault detection, testing of physical theories and evaluating the predictability of time series.

The range of values taken by the samples of a signal and how often the signal assumes these values together determine the signal variability. The signal variability can be seen by plotting the time series and is quantified by the histogram of the signal samples, which shows the percentage of the signal amplitude values with in a certain range. The numerical description of signal variability, which depends only on the value of the signal samples and not on their ordering, involves quantities such as mean value, median, variance and dynamic range.

B. *Correlation*

Scatter plots give the existence of correlation, to obtain quantitative information about the correlation structure of a time series x (n) with zero mean value.

The better estimate is to use the empirical normalized autocorrelation sequence, which is an estimate of the theoretical normalized autocorrelation sequence. For lag L = 0, the sequence is perfectly correlated with itself and we get the maximum value of 1. If the sequence does not change significantly from sample to sample, the correlation of the sequence with its shifted copies, though diminished, is still close to 1. Usually the correlation decreases as the lag increases because distant samples become less and less dependent. Signals, whose empirical autocorrelation decays fast, such as exponential, have short memory or short-range dependence. If the empirical autocorrelation decays very slowly, as a hyperbolic function does, then the signal has a long memory or long-range dependence.

The spectral density function shows the distribution of signal power or energy as a function of frequency. The autocorrelation and the spectral density of a signal form a Fourier transform pair and hence contain the same information. However, they present this information in different forms, and one can reveal information that cannot be easily extracted from the other.

Various tools found in the literature [3][4][5][6] which are used in the classification of signals are:

- Power spectrum
- Correlation dimension
- Embedding parameters
- Lyapunov exponent
- Surrogate data study
- Deterministic Versus Stochastic plot
- Structure function
- Average Mutual Information
- False Nearest Neighbors
- Embedding dimension
- Hurst exponent
- Recurrence Histogram
- Spatio-Temporal Entropy

III. Experimental Results

The data taken (1447 samples) for this experimentation are

- ❖ BSE index for the years 1999 – 2004
- ❖ NSE index for the years 1999 – 2004

A. *Time Series Chart*

Time series chart is the regular 2-dimensional graph that shows the dynamics of the scalar series in time. Time series chart are shown in Figure 2 and 3.

B. *Auto Correlation Function*

The auto correlation functions for BSE and NSE index are shown in Figure 4 and 5. This shows long range dependence (LRD) for both NSE and BSE.

C. Correlation between NSE and BSE

Pearson coefficient r, gives a measure of association between two variables.

r = 1 perfect Positive relationship
r = 0 No relationship
r = -1 perfect negative relationship

Relationship between NSE and BSE for the period is explored and the results are shown in Table I.

A total of 1447 sample data was taken for the experiment. The correlation is significant at the 0.01 level (2-tailed).