

## **The Evolving Importance of Statistics in Securities Market**

Sarat Malik and Suvidha Nagpal

*SEBI, Mumbai*

Final Version Received on 30.08.2017

---

### **Abstract**

Big, Unstructured, heterogeneous and temporal data is being generated every second in the stock market and it requires a new school of thought which can not only handles its complexities but also be able to help in the future prediction and analytics of the market. Knowledge discovery and future forecasting will not be possible without handling the core challenges of big data. Stock market is one of the burning areas where data is growing day by day. Due to the the heterogeneity and other complexities of data, a structured data architecture and design is needed which specifically deals with the stock market data and analyze the heterogeneous data for the future prediction of the market. Also, big data analytics is a very promising area and buzz word for the next generation information technologies.

---

### **1. Introduction and Review**

Human reasoning is based on the decision maker's ability to process what they know, whether they learn by example or create new approaches. Reasoning is used by managers to discover patterns in data, which helps to infer multiple meanings from a single input and generalize from diverse inputs. Intelligent information systems can aid and enhance decision making by applying specific reasoning tools: model-based reasoning, abstract reasoning, heuristic reasoning, procedural reasoning, analogical reasoning and case-based reasoning [3]. As we enter the new millennium, innovations and competitive pressures force organizations to rethink the manner in which they conduct business and to redefine the essential ingredients for success [4]. The basis of competition is being fundamentally altered through the introduction of the Internet and other advanced technologies (Sampler, 1998) [10].

The securities market corresponds to the financial system segment that involves operations on securities usually tradable on financial markets, either organised or non-organised. This market includes, on the one hand, a short-term component where securities such as Treasury Bills or Commercial Paper are included and, on the other hand, a long-term segment where debt securities, such as bonds only shares and other equity instruments are traded. The securities market may also be viewed from two perspectives: the primary market which includes financial assets that enter the market through issues, and the secondary market that comprises financial assets already in circulation and subject to transactions in or out of stock exchanges. The development of a nation's productive capacity requires capital formation which can be done either by utilizing domestic resources or through external assistance and securities market serves as an important medium for channelizing domestic savings into investments.

---

Views expressed in this article are those of the author only and not of the organization to which he belongs.

Corresponding Author: Sarat Malik

E-mail: saratm@sebi.gov.in

The importance of securities markets in intermediating financial flows, both domestically and internationally, underscores the need for relevant, coherent, and internationally comparable statistics. This need was recognized by the G-20 Data Gaps Initiative, launched in the aftermath of the 2008 global financial crisis with the support of the G-20 Finance Ministers and Central Bank Governors and the International Monetary Fund's International Monetary and Financial Committee. Good securities data, along with monetary and financial statistics, are critical to understand the evolution and structure of financial intermediation. Economic policy decision-makers, and in particular central banks, also felt the need for new statistical information at the securities market level. Securities statistics are recognised not only as a useful autonomous instrument in the analysis of financial markets, but are also used as information sources in monetary and financial statistics, balance of payments and international investment position statistics and national financial accounts. It is a well-known fact that capital markets provide impetus for economic growth. A well-functioning capital market helps the growth process in an economy by increasing savings, making increased savings available for investment, allocating investment resources efficiently and ensuring better utilization of the existing resources. A variety of saving investments in the form of Mutual Funds, Private Equity, Municipal Bonds, Reits & Invits, AIFs etc are available.

Research finds that both the operation of banks and the functioning of securities markets influence economic development (Demircuc-Kunt and Maksimovic, 1998; Levine and Zervos, 1998), suggesting that banks provide different services to the economy from those provided by securities markets. The securities market facilitates the internationalisation of an economy by linking it with the rest of the world. This linkage assists through the inflow of capital in the form of portfolio investment. Moreover, a strong domestic stock market performance forms the basis for well performing domestic corporate sector to raise capital in the international market.

A stock exchange is an entity which provides "trading" facilities for stock brokers and traders to trade stocks and other securities. How to invest in stock exchange is one of the important issues in investment, and one of the factors that can help investors in the process of investment is the efficiency of the corporation under consideration. In traditional point of view, the most important objective of efficiency appraisal was a review of the historical efficiency [5]. Although big data concepts in the financial market are relatively new and are introduced in the last few years, internet and related technologies are not new for it. The realities of the electronic marketplace make Internet technology the most used form of IT (Ferguson, 1996) [2]. Stock investment has been a major investment tool for personal wealth. Stock investors have been eager to know how to select stocks to gain profit. Generally speaking, there are two most often used analytical methods when one selects stocks to invest: fundamental analysis and technical analysis. Fundamental analysis mainly focuses on the listed companies' operation and financial status to forecast future profit/loss and one can then select stocks accordingly. Technical analysis focuses on historical stock price movement, which might show a pattern or trend used by investors to forecast the future possibility of the rise/fall of the price, and the investor can then decide whether to invest or not [1]. Researchers are continuously trying to explore new methodologies for the stock market prediction. Volume of stock prices traded per session is crucial in the forecasting methodologies [6]. One of the original works describing the application of evolutionary algorithms to stock trading can be found in Oussaide`ne et al. (1996, 1997) [8] [9]. The usefulness of a model of this type depends upon its ability to forecast future changes in stock market valuations [7]. There is dearth of literature available on the big data for the stock market. An ample amount of research evidence is available for the big data in financial

market but there is hardly any evidence of customized big data architecture over Hadoop for the stock market.

## 2. Importance of statistics

Data is fast becoming the “fourth organizational pillar” for modern financial institutions, alongside people, processes and technology. In today’s financial and economic environment there is an urgent need for the securities market regulator to make well-informed policy decisions. Data management has been pushed to the forefront today in order to make policy decisions, compliance checks, risk management, cost efficiency of operations, benefits for investors, and research by academicians. All of these functions rely on the accuracy and integrity of data for effective decision making and policy formulation. In this modern era of technology, we are witnessing a wave of technological advancements. These advancements are as diverse in the areas like collection, storage, aggregation, processing and analysis of financial data. Moreover, with the advent of the Internet of things, there is an exponential rise in huge volumes of unstructured data from several different sources.

The data to be used can be either collected using data gathering techniques or someone else’s existing data, if it serves the purpose of the research. Collecting the data correctly takes a great deal of work. Before data analysis can begin, the accuracy of the data collected needs to be verified. Following data collection, the data needs to be critically analysed. For any research, data analysis is very important as it provides an explanation of various concepts, theories, frameworks and methods used. It eventually helps in arriving at a conclusion and proving the hypothesis.

**Data vs. Information:** Data are plain facts. When data is processed, organized, structured or presented in a given context so as to make them useful, it is called information. It is not enough to have data. Data in themselves are fairly inadequate. But when these data are interpreted and processed to determine their true meaning, they become useful and can be called information. Data is the computer’s language. Information is our translation of this language.

## 3. Nature and Behavior of Data in Stock Market

According to the data from Securities and Exchange Board of India (SEBI), there are more than 85,000 intermediaries registered in India and it is growing day by day. Intermediaries are developing their own repositories and due to it, same data has been saved at all places. It is using power, space, energy and cost. In current market scenario, future forecasting requires mammoth of data coming from every corners of the globe. Data repetition at all databases of interested parties is the great concern for green computing practitioners. Indian as well as world markets have witnessed many downfalls in the past. Mayhem in the financial market downgrades the confidence of the middle class population. In year 2008, subprime lending in US had impacted the world markets and a global recession or slowdown was witnessed. Similarly in 2008–09, information about Satyam created a vacuum in the market and small investor’s committed suicide. Little and slow access of information became big reason for the huge loss in the market for small investors. All information need to be integrated quicky and provided to the small and medium investors at very reasonable and affordable cost. Traditional information database will not be able to handle and integrate this type of heterogeneous, temporal, unstructured and dynamic data stored at different places, so a new architecture is required for the integration of the heterogeneous database. Data can’t be available at one place or one database and it is in the

different format. It requires an integrated mechanism which provides all temporal data at one place and easy to process for the decision making process in desired format.

Stock market is affected by all information's. As per the efficient market hypothesis, stock market must react on all of the information. Small and Middle class investors do not have such type of resources to know all happenings instantaneously and take their positions respectively. Easy availability of this integrated data from all corners of the globe is the key issue. Due to temporal nature of the data required in the stock market, big data architecture is being implemented in such a way that all data must be updated and integrated on timely basis and reports should be generated based on the current information. To achieve efficient market hypothesis, this information must reach all participates of the market. Due to shared resources and other above discussed benefits of big data technologies, small investors too have equal opportunity to excess relevant data. This will increase participation, participation will increase more investments, investments will increase more stable market with market hypothesis, more realistic market hypothesis will lead to more accurate prediction of the market and finally it leads to stable and mature market. These will create investment cycle as shown in figure below [6].



Participation will increase investment in the market. Investment will increase fractal behavior and that leads to more accurate prediction of the market. Final this process will move forward towards more stable market in the third world countries like India also. Behavior of the stock market will entirely change after implementation of the big data technologies in the stock market.

#### **4. Maintenance of Statistics at SEBI**

The research function of SEBI is delineated under Section 11(2) (1) of the SEBI Act, 1992. Economic research contributes to the development and updating of conceptual models and tools which form the basis for its analysis and policy recommendations. Conducting high-quality, policy-relevant research for a regulator is essential for its credibility both in interactions with country authorities and with the international community in general. Research activities help facilitate understanding of the multiple economic uncertainties that lie ahead. Sound research backed by relevant data assist any policymaking body in carrying out its activities under increasingly challenging times. Availability of strong financial data gives the regulator the ability to make faster, more informed policy decisions which are backed up by facts which further leads to deeper understanding amongst the investors, market and the regulator.

Financial markets are one of many areas which have big appetite for data. In securities market we use analytics predominantly for key functions such as trading, compliance, and risk management. Market intermediaries like exchanges, clearing houses, depositories, and the regulators too have known the importance of analytics. The capital markets industry is one of the most data driven industries. Electronic trading generates millions of messages every day. Regulatory and risk management requirements are challenging banks and financial services firms to capture, store, and analyse data that spans multiple years, departments and region. The amount of transactional data is growing at an exponential pace due to the ever increasing volume of transactions.

**Periodicity of Data:** Securities and Exchange Board of India publishes a detailed report annually which provides a full and true account of the activities undertaken by SEBI during a particular financial year. The report covers data related to securities markets and is present in comparable format with reference to previous years. Besides the annual report, quarterly reports detailing all the policy developments and data/information pertaining to Indian securities market for each quarter is submitted for information at the SEBI Board meetings.

Monthly Bulletin encapsulates all the regulatory developments for the month and aggregating the data for the securities market. On a similar line securities market data is available on weekly and daily basis. Data such as total number of registered market intermediaries, market capitalization, turnover, depository statistics, only assets under management (AUM), FPIs and macroeconomic indicators are readily available for exploration and analysis.

Under SEBI's Development Research Group (DRG) wide-ranging data sets are required to carry out extensive policy oriented research to examine the existing policies from an academic perspective as well as to suggest new policy alternatives, essentially on subjects contributing clarity and solutions to challenges faced by SEBI.

#### **4.1 Types of Data in securities market:**

- Mutual Funds – Net Inflow/Outflow, Assets under management (AUM), AMFI
- Foreign Portfolio Investors- Equity, Debt
- Fund raising from primary market- Public issues, Right issues, private placement
- Secondary market- market returns, market capitalization, turnover, trading volume,
- Indices: Domestic, International, Volatility, PE Ratio
- Corporate Bonds
- Alternative Investment Funds (AIFs)
- Macroeconomic and Financial Indicators- GDP, CPI, WPI, Exports, Imports, BoP
- Depositories: DP accounts, Demat value, Demat quantity,

#### **5. Uses of Data**

The users of the capital market data are widespread. They include the Regulators, Policy makers, Government, Investors, Stock exchanges, Depositories, Clearing Corporation, academicians, and researchers amongst others. SEBI is also responsible for providing regular data to the Ministry of Finance (MoF), the Reserve Bank of India (RBI), Ministry of Corporate Affairs and Government of Maharashtra for their frequent updation and for supporting informed policy decisions. Data support includes contributions for preparing GoI's Economic Survey, the Government of Maharashtra Economic Survey, the IMF Redbook, Statistical Commission and the Mid-year Economic Analysis.

Financial reporting is used by Regulators, investors, institutions, government, and others to make important decisions. Under the disclosure norms the listed entity shall submit quarterly and year-to-date standalone financial data to the stock exchange within forty-five days of end of each quarter, along with Limited Review Report or Audit Report as applicable. Prospective investors make use of financial data to assess the viability of investing in a business. Data facilitates the functioning of SEBI and helps in quick and easy decision making in the dynamic market environment.

## 6. Big Data Analytics

Big data analytics helps organizations harness their data and use it to identify new opportunities. That, in turn, leads to smarter business moves, more efficient operations, and higher profits. What makes Big Data different from other data groups is the vastness and range of data that is available.

**Data gap:** systematic attempt to capture securities market data have to be made

- Lack of proper format as compared with the available banking sector data real sector data
- Lack of consolidated data in securities market
- No single agency collects and disseminates data to public. Data is fragmented.
- No serious attempt by the Government to collect and patronize the securities market data.

**Different users draw different insights:** two different users confronting the same data may not necessarily draw the same conclusion, depending on their previous experiences and particular level of expertise.

## 7. Concluding Remarks

Big data integration will have a great impact on the growing markets like India. There is no need to store same data again and again. Intermediaries and other interested parties will be benefited in two ways. First, they can get data without wasting time and cost as per their own desire format. Secondly, they can exploit other existing benefits of Big or cloud computing like SaaS, Paas, IaaS etc. Implementation of big data technologies and above design architecture will reduce use of resources while maintaining same efficiency and effectiveness. Implementation of green computing will also helpful in the forecasting technologies or methods.

Indian stock markets are still working with the traditional database and only upper class of investors has capacity to afford information from all database maintained by different agencies. Because of it, market theories and definitions like efficient market hypothesis, fractal market hypothesis etc are not applied properly in the third world markets. Extensive, heterogeneous, unstructured and temporal data required for the market prediction will be handled only by big data technologies, i.e. cloud computing, parallel computing, high performance computing, distributed computing etc. Unique feature of cloud computing like sharing of resources must be incorporated in the market for more sustainability of the investment process and economies. In spite of all these benefits there are few hiccups also which must be dealt with. Few of them are:

1. Ownership of big data is still not defined and it is big hurdle in the framework design and implementation of such architecture.
2. Security issues of the data exist.
3. Data redundancy at various places by intermediaries is the greater concern for the greener technologies.

**References**

- Chien-Jen, Huang, Peng-Wen, Chen and Wen-Tsao Pan. (2011). *Using multi-stage data mining technique to build forecast model for Taiwan stocks*. Springer-Verlag London Limited.
- Ferguson, G.T. (1996). Strategy in the digital age: Role of information technology in corporate strategic planning. *Journal of Business Strategy*, **17**(6), 28–31.
- Hand, D.J. (2008). Mining the past to determine the future: problems and possibilities. *International Journal of Forecasting*, **25**, 441–451.
- Hoplin, H.P. (1995). Re-engineering information technology: An enabler for the new business strategy. *Industrial Management & Data System*, **95**(2), 24–27.
- Kamran Rezaie, Vahid Majazi Dalfard, Loghman HatamiShirkouhi and Salman Nazari-Shirkouhi. (2013). Efficiency appraisal and ranking of decision-making units using data envelopment analysis in fuzzy environment: a case study of Tehran stock exchange. *Neural Comput & Applic*, **23**, 1-17. DOI 10.1007/s00521-012-1209-6.
- Krishna, Kumar Singh, Dr. Priti, Dimri and Madhu, Rawat (2013). Fractal Market Hypothesis in Indian Stock Market. *IJARCSSE*, **3**(11), 739-743.
- Patrick, Thomas. (2001). A relationship between technology indicators and stock market performance. Jointly published by Kluwer Academic Publishers, Dordrecht Scientometrics, and Akadémiai Kiadó, Budapest, **51**(1), 319–333.
- Oussaide`ne M., Chopard, B., Pictet, O.V. and Tomassini, M. (1996). Parallel genetic programming: an application to trading models evolution. In: *Proceedings of the first annual conference on genetic programming*, GECCO '96. MIT Press, Cambridge, 357–362.
- Oussaide`ne, M., Chopard, B., Pictet, O.V. and Tomassini, M. (1997). Parallel genetic programming and its application to trading model induction. *Parallel Comput*, **23**, 1183–1198.
- Sampler, J.L. (1998). Redefining industry structure for the information age. *Strategic Management Journal*, **19**(4), 343–355.