

Does Big Data Influence the Efficiency of the Capital Markets?

Rajesh Kumar Singh, S. K. Mitra & Sumeet Gupta

This paper examines the adaptation of the 'big data' strategies in the developed capital markets and its effect on the efficiency of the capital markets. The big data strategy and algorithms use the power of high capacity computing to affect the high frequency trading which improves the efficiency in the market. However, high frequency trading also poses many regulatory challenges for the Security and Exchange Commission. Social media and microblogs affect the risk appetite of the investors. The sentiment and decision-making pattern of the investors are influenced by the continuous flows of the information through the social media which affects the capital markets.

Rajesh Kumar Singh (E-Mail: rajesh.efpm2013@iimraipur.ac.in), **S. K. Mitra** (E-Mail: skm@iimraipur.ac.in) & **Sumeet Gupta** (E-Mail: sumeetgupta@iimraipur.ac.in) are from Indian Institute of Management Raipur

Introduction

Within the financial services sector, 'big data' has gained far more traction within retail banking and insurance sectors due to the increasing desire of these financial institutions to profile their customers in a similar manner to early adopters of big data strategy such as Amazon, Baidu or Google. Research by the International Data Corporation indicates that the global data volume is expected to reach 35 zettabytes (ZB)¹ by 2020. Beyond that, the trend of growth of doubling the data every 2 years will be maintained. This implies that we have entered the era of big data. On the institutional side of the capital markets, there had traditionally been far more customer stickiness, hence there had been less incentive to apply big data in this manner.

Big data is defined as, "data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data" (Snijders et.al, 2012). However, big data strategies had begun to make some impact in a se-

¹ In computer terminology, ZB refers to one sextillion bytes

lected few areas of the capital markets including the sentiment analysis for trading and growth in volume, risk analytics, fraud prevention and market surveillance. The business landscape has been going through constant change since the financial crisis in 2008. The emergence of internet and social media networking services combined with the extensive dissemination of smart phones has revolutionized the way we communicate and exchange information. Most of the firms have been using the information obtained from the vast oceans of available structured and unstructured data to gain customer knowledge, anticipate market conditions, and better gauge customer preferences and behavior ahead of time so that firms could offer highly personalized customer-centric products and services such as sentiment analysis-enabled brand strategy management and real-time location-based product offerings as opposed to the historically offered product-centric services. The global financial markets had increasingly been fragmented due to rapid globalization and technological changes (Funk et.al, 2014). In one of his studies on liquidity, Blocher et.al (2016) identified three key components of the financial market like 1) fund management for long term investors 2) low-frequency trading (LFT) by the traditional brokers and 3) high-frequency trading (HFT) by proprietary financial firms which used the big data and ag-

The global financial markets had increasingly been fragmented due to rapid globalization and technological changes.

gregated information to set the trading strategy. Continuous flow of the big data from increasing online activities by users has become a buzz word for recent years because of its potentials for various uses including marketing, trading, political predictions, disease epidemics, social dynamics, etc. This is the era of information explosion and a world overwhelmed by numbers and digits. Also the market events, technology changes and regulatory reforms keep on throwing new challenges for the capital markets. Solution to many of these issues which capital market faces lies within house by analyzing and leveraging the huge volume of the data accumulated on day to day basis. The capital market industry has varied data sources, which included structured data like traditional banking transaction data and market data. At the same time, it also generates mammoth volume of unstructured data through corporate news, feeds, macro and micro economic indicators, social media updates and contents. In recent years, capital markets have gone through an unprecedented change, resulting in the generation of massive amounts of high-velocity and heterogeneous data. Similar trends could be observed in the financial services sector as well, where big data has been increasingly becoming the most significant, promising, and differentiating asset for the financial services enterprises (Seth et.al, 2015). These massive data troves could be processed through big data strategy, tools and techniques which could be game changer. Traditional tools cannot process such large datasets, however the Big Data based approach can analyze structured and un-structured data and

create logical patterns to help business to take decisions. The speed and agility of this processing is exponentially faster than that could be done by the traditional mechanism. This provides near real time information with the actionable intelligence which could be used in the decision making process. The analytical and predictive power of information generated from online big data for the capital market activity is supported by numerous studies ranging from stock markets to housing markets.

Big data strategy could help the capital market to address key use cases in the area of: 1) trading strategy, 2) reporting, 3) compliance and 4) operational simplifications. Investment banks have an untapped opportunity to use big data to solve many of their business problems. The age of big data offers new creative opportunities across the board. For finance researchers, big data allows the field to settle old debates and discover new puzzles. For investment practitioners, the big data approach offers limitless possibilities to gain informational advantage over the competition by cleverly analyzing public data sources. Big data strategies in the capital markets tend to be synonymous with analytical tasks or those related to reporting or governance functions but in recent years the consumption of text-based, audio and video unstructured data had also been a significant driver for some projects. The big data story in the capital markets is directly tied to the rising importance of data management as a function within financial institutions. Regulatory, client, and internal drivers have forced most

firms to re-evaluate the core reference data sets on which they are basing their trading, risk management, and operational decisions.

Capital Markets & Technologies

Big data had been a much misused and misunderstood term within the financial services industry and capital markets for some time, applied to everything from traditional relational databases (RDBMS) to web-based sentiment analysis tools. A data driven culture is driving the research field in finance and management since the advent of structured, semi-structured and unstructured socio-economic and demographic information from social media, mobile devices, blogs and microblogs. Big data, which encompasses this revolution, involves the usage of new tools for financial professionals and academic researchers due to the size of data involved. In the capital market, the technology focus is often towards improving real-time systems, transaction processing, operational efficiency and establishing data warehouses of structured data. The adoption of discontinuous or disruptive technologies, there is a gap or chasm between innovators and early adopters and that is where

A data driven culture is driving the research field in finance and management since the advent of structured, semi-structured and unstructured socio-economic and demographic information from social media, mobile devices, blogs and microblogs.

big data is today in capital markets. The capital markets tend to deal largely in structured data sets from a more limited set of sources-market data vendors, market infrastructures, and counterparties. Some unstructured data sets have however, become important to capital market institutions in areas such as sentiment analysis and market surveillance for profiling certain trends and activities in the market and within firms, but these have not traditionally been the data sets of primary importance to business. The big data refers to data that could be categorized in the four categories: 1) volume and quantity of the data generated, 2) velocity and speed at which data are generated, 3) variety and types of data like text, audio and video etc and 4) veracity and integrity and quality of the data being processed. In the context of big data, it was represented by 4 vs (variety, velocity, veracity and volume) but while understanding its context in terms of capital market, it is important to consider the other attributes of the data such as variability, value and visualization. These 7 vs made the complete view of the data in the context of capital market. The large scale enterprise level Natural Language Processing (NLPs) affect the capital market strategy for the High-Frequency Trading (HFTs) and Low-Frequency Trading (LFTs). These new emerging technology and cutting edge technologies are bound to have significant impact on the capital markets in the developed and emerging economies. Big data platform has been in the forefront to bring innovation in the areas of Data Science, Artificial Intelligence, Machine Learning, Deep Learning and Natural

Language Processing (NLP). At the verge of big data crossing Moore's chasm as the capital markets have started focusing on the unstructured data for new avenues for innovation and has offered immense opportunities for growth and sustainability.

Big Data & High Frequency Trading

High frequency trading in the developed capital markets is a very recent phenomenon using the high volume of the enterprise data and continuous flow of the information and content from the social media for analysis, it came to general public attention on July 23, 2009 in a New York Times article (Duhigg July 23, 2009)². High frequency trading (HFT) is one of the forms of trading where control is taken over by the high capacity computers using proprietary artificially trained and machine learned algorithms. These algorithms are generally developed on the top of the empirical data trained to the machines based upon the models developed by Data Scientist. Big data and analytics based computing could benefit today's banking and financial organizations on many aspects, and provide much valuable information for organizations to achieve more intelligent trading in the capital market, which could help them to gain a great competitive advantage. However, the large scale of data and the critical latency analytics requirement in finance poses a great challenge for current system architecture (X Tian

² <http://www.nytimes.com/2009/07/24/business/24trading.html> accessed on 22nd February 2018

et.al, 2015). Industry has started looking at the capital markets through the lens of Big Data platform using the innovations in the areas of Data Science, Artificial Intelligence, Machine Learning, Deep Learning and Natural Language Processing (NLP). The big data approach along with the machine leaning and NLP could affect the empirical data processing in the high frequency trading (HFT) in the capital market. These new emerging technologies are bound to have significant impact on the capital markets in the developed and emerging economies. In the developed capital markets like USA, Japan and Europe, algorithmic high frequency trading is quite advanced. One study by Nagata et.al (2014) in the context of Nikkei stock exchange observed that high frequency trading improves the efficiency of the market but it also observed that it was disadvantageous for those customers not having very refined trading algorithms and high speed internet services. One of the researches in the context of the capital market based upon the empirical data collected from HFT firms and regulators in the US and UK to study the influence of data analytics in the high frequency trading (Jonathan et.al, 2017) observed competitive asymmetries between HFTs and low-frequency traders (LFTs) operating in the more traditional forms of market trading. These findings produced significant evidence that HFT gained

HFT gained extensive market advantages over LFT due to significant investment in advanced computing and technologies.

extensive market advantages over LFT due to significant investment in advanced computing and technologies. To keep the capital market transparent and remove the competitive asymmetry, regulators are challenged to keep pace with the evolution of HFT based upon the newer technologies platforms. In one of the studies by Shorter and Miller (2014), it is observed that in 2012, algorithmic trade instructions sent by both LFT and HFT accounted for over 1.6 billion shares every day. The technical relationship between big data and HFT coupled with the customized trading software based algorithms helped to define how judgments could be made in the financial markets (Angel & McCabe, 2013). Taking advantage of the high computing power, high frequency traders could develop ability to revise their quotes quickly after any positive or negative news which could lead to the reduction of inefficiency of the capital market. In one of the studies on the market microstructure and associated risk of high frequency trading, Aldridge et.al (2013) observed that high frequency trading reduces the arbitrage due to minimized gap in information.

Ait-Sahalia and Saglam (2013) observed that data asymmetries which created disadvantage for the other traders were exploited by the HFTs for their advantage. For instance, about 70% of the US equity trades today are generated by high-frequency trades (HFTs) and are machine driven. In a recent study Blocher et al. (2016) observed at NASDAQ for the S&P 500 that the high levels of order cancellations were essential for HFT firms to establish a true

market price. As it is always understood that speed is the defining attribute that sets high-frequency trading (HFT) apart but it is the natural evolution of a new trading paradigm that is characterized by strategic decisions made in a volume-clock metric. It is imperative to understand what will happen if the speed advantage disappears in future, but it was observed by Easley et.al (2012) that HFT will still evolve to continue exploiting structural weaknesses of low-frequency trading (LFT).

Big Data & Capital Market Efficiency

In the world of finance, the efficient market hypothesis (EMH) asserts that financial markets were 'informationally efficient' which means current stock prices were already reflecting all known information (structured or unstructured) and all occurred events and facts. Therefore, investors cannot make excess profits from the market if their trading strategies are based on known information, because market prices are efficiently collecting and aggregating information and keep changing without delay (Zhang & Skiena, 2010). Enterprise headquarters faces a tradeoff between the cost of attaining accurate private information and the value of the information which could lead to the inefficacy in the capital market. Foster (1979) studied the efficiency and inefficiency of the capital market and explained the influence of information on the capital market. Fama (1991) made the simple hypothesis that security prices fully reflect all available information considering the precondition

of that information and trading costs, the costs of getting prices to reflect information, are always zero. In the case of competitive equilibrium, all arbitrage profits are eliminated, it is not possible to keep the competitive economy always be in equilibrium state. It is clear that those who arbitrage make no return from their costly activity in case market is ideally information efficient. Hence the assumptions that all markets, including that for information, are always in equilibrium and always perfectly arbitrated are inconsistent. In the advent of structured and unstructured information flow in the capital market through the social media, the role of information becomes even more relevant. In case the access of information with all the actors in the capital market not in equilibrium, then it leads the market toward arbitrage and makes it less efficient. However, effective usage of the big data information could reduce the opportunity of arbitrage and make it more efficient. Fama (1991) was also of the opinion that predictability of the stock returns based upon the dividend yields is not in itself evidence for or against market efficiency. While understanding the behavior of the capital market, Sharpe (1964) mentioned that to predict the behavior of capital markets in the state of equilibrium, capital asset prices have adjusted so that the investor, if he follows rational procedures through diversification is able to attain desired point along a

Effective usage of the big data information could reduce the opportunity of arbitrage and make it more efficient.

Table 1 Findings of Key Papers

S. No.	Year	Authors	Paper title	Does Big Data in fluence the efficiency of the Capital Markets	Major findings
1	2017	Jonathan et.al	A model for unpacking big data analytics in high-frequency trading	Accepted	Big data influenced the HFT significantly and caused competitive asymmetry between HFT and LFT
2	2017	Ye, et.al	Internet big data and capital markets: a literature review	Accepted	The continuous flow of the information influences the investors decisions in the capital markets
3	2016	Blocher et.al	Phantom Liquidity and High-Frequency Quoting	Accepted	High-frequency trading firms behave as agents who could bring efficiency to the capital market.
4	2016	Bartov et.al	Can Twitter help predict firm-level earnings and stock returns?	Accepted	Observed positive association between the aggregate opinion and the immediate abnormal stock price reaction to the quarterly earnings announcement
5	2016	Jin et.al	Has microblogging changed stock market behavior? Evidence from China	Accepted	Observed significant impact of microblogging to increase the relative trading volume as well as the decreases in the daily expected stock return
6	2016	Singh, A. P.	Do Technology Spillovers Accelerate Performance of Firms? Unravelling a Puzzle from Indian Manufacturing Industry	Accepted	Technology influences productivity
7	2016	Singh, A. P.	R&D Spillovers & Productivity Growth: Evidence from Indian Manufacturing	Accepted	R&D promote the adoption of new technologies such as big Data and Machine Learning
8	2015	T Seth et.al	Big Data: Algorithms, Analytics, and Applications	Accepted	Big Data in finance will evolve to influence the trading the pattern of the investors

9	2015	X Tian et.al	Latency critical Big data computing in finance	Accepted	Big data and Analytics based computing could benefit today's banking and financial organizations to achieve more intelligent trading in the capital market
10	2014	G Shorter et.al	High-frequency trading: background, concerns, and regulatory developments	Accepted	High frequency trading will keep on posing the regulatory challenges with the development of technology
11	2014	Funk et.al	Derivatives and deregulation: Financial innovation and the demise of Glass-Steagall	Accepted	Big Data implications on the HFT trading is highly tilted due to technology
12	2014	Nagata et.al	Does High-Speed Trading Enhance Market Efficiency? Empirical Analysis	Accepted	High efficiency trading enhances marketing efficiency
13	2014	Moat et.al	"Arrowhead" of the Tokyo Stock Exchange Using big data to predict collective behavior in the real world	Accepted	Observed the predictive effect of the Big Data analysis in the financial world
14	2013	Aldridge	Market microstructure and the risks of high-frequency trading	Accepted	High frequency trading reduces the arbitrage due to gap in information
15	2013	Angel et.al	Fairness in financial markets: The case of high frequency trading	Accepted	Impact of High frequency trading was put into question as the fair practice in the capital market.
16	2013	Aït-Sahalia et.al	High frequency traders: Taking advantage of speed	Accepted	High Frequency traders could take the advantage of the speed to influence the capital market in their favour
17	2013	Alanyali et.al	Quantifying the relationship between financial news and the stock market.	Accepted	Observed that movements in financial markets and movements in financial news are intrinsically interlinked
18	2012	Snijders et.al	Big Data": big gaps of knowledge in the field of internet science	Accepted	Observed that Big Data could bridge the gap in knowledge in the

19	2012	Easley et.al	The volume clock: Insights into the high-frequency paradigm	Accepted	capital markets and reduce efficiency
20	2011	Bollen et.al	Twitter mood predicts the stock market	Accepted	High frequency trading has insightful impact on the efficiency of the capital market Behavioral economics tells that emotions can profoundly affect individual behavior and decision-making
21	2011	Zhang, et.al	Predicting stock market indicators through twitter "I hope it is not as bad as I fear"	Accepted	Emotional tweets percentage is significantly negatively correlated with key indices in the USA market.
22	2010	Zhang et.al	Trading Strategies to Exploit Blog and News Sentiment.	Accepted	Large scale natural language processing (NLP) system did affect the capital market strategy for the HFT and LFTs
23	2010	Gilbert et.al	Widespread Worry and the Stock Market	Accepted	Our emotional state influences our choices even in the capital markets
24	1991	Fama, .	Efficient capital markets: II		Security prices fully reflect all available information
25	1979	Foster	Briloff and the capital market	Accepted	Influence of the information on the Capital market efficiency and inefficiency was observed.
26	1964	Sharpe	Capital asset prices: A theory of market equilibrium under conditions of risk.		Capital assets prices achieve equilibrium near the Capital market line (CML)

capital market line. Investor should be able to absorb all the available information and adjust and diversify its portfolio so that he could attain the desired goal along the Capital Market Line (CML).

It was difficult to observe the investors views on the capital markets and understand the decision making process before the advent of big data. With the ubiquity of information technology and the Internet, an increasing number of investors are gathering information from the Internet for analysis. Ye et.al (2017) observed that Google processed 2 million searches per minute in 2012, 204166667 emails were sent per minute, Twitter users posted approximately 100,000 tweets per minute pertaining to the capital markets. Most of the data related to capital markets is categorized into three categories: 1) forum type, 2) micro blog type and 3) internet search type. Bartov et.al (2016) aggregated the opinion from individual tweets to successfully predict a firm's forthcoming quarterly earnings and announcement returns. In order to study the impact of microblogging Jin et.al (2016) observed a significant increase in the relative trading volume as well as decrease in the daily expected stock return and firm-level volatility. In similar lines, while studying the impact of the financial news in the media Alanyali et.al (2013) observed intrinsic impact on the movements of the financial markets. While big data is spreading, a very convincing evidence is the impact observed by Moat et.al (2014), which infers that online data and information can give new insights into the real-world which could affect the collective decision making and

can even anticipate future actions in the capital markets. In the field of behavioral economics Bollen et.al (2011) observed that emotions can profoundly affect the individual behavior and decision-making which could even be extended to prediction of economic indicators. Zhang et.al (2010) extended the research even further and observed based upon the analysis of the Twitter feed on daily basis and its correlation with the indices (NASDAQ, S&P 500, and Dow Jones), that emotional tweets percentage is significantly negatively correlated with key indices in the USA market. This is significant revelation which indicates the impact of the microblogs on the capital market of the developed countries like USA. When people are pessimistic or uncertain about the future, they will be more cautious to invest and trade. So capturing the collective mind – especially people's mood becomes one possible way to predict the stock market movement. While studying the human emotion and its influence on the capital market Gilbert et. al (2010) observed that human state of emotion can influence decisions even in the capital market and the investment.

Concluding Remarks

While all the studies in the past related to impact of big data and its influence on the capital markets indicate that information asymmetry could reduce to a significant level. However many other macro variables like availability of internet to all the investors, speed, access free of cost to relevant information and biasness in the content on the social

media could provide opportunity for inefficiency in the market? Capital markets will evolve and use the continuous flow of information through defined and non-defined sources such as social media, microblogs as part of their big data strategy to do the sentiment analysis and improve the decision moments to minimize the inefficiency in the capital market.

References

- Aldridge, I. (2013), "Market Microstructure and the Risks of High-frequency Trading", <http://dx.doi.org/10.2139/ssrn.2294526>
- Ait-Sahalia, Y. & Saglam, M. (2013), High Frequency Traders: Taking Advantage of Speed (No. w19531), National Bureau of Economic Research
- Angel, J. J. & McCabe, D. (2013), "Fairness in Financial Markets: The Case of High Frequency Trading", *Journal of Business Ethics*, 112(4): 585-95
- Alanyali, M., Moat, H. S. & Preis, T. (2013), "Quantifying the Relationship between Financial News and the Stock Market", *Scientific Reports*, 3: 3578
- Blocher, J., Cooper, R., Seddon, J. & Vliet, B. V. (2016), "Phantom Liquidity and High-Frequency Quoting", *The Journal of Trading*, 11(3): 6-15
- Bollen, J., Mao, H. & Zeng, X. (2011), "Twitter Mood Predicts the Stock Market", *Journal of Computational Science*, 2(1): 1-8.
- Bartov, E., Faurel, L., & Mohanram, P. (2015), "Can Twitter Help Predict Firm-Level Earnings and Stock Returns?" *The Accounting Review (In-Press)*, <http://aaajournals.org/doi/10.2308/accr-51865> accessed on February 24, 2018
- Easley, D., López de Prado, M. M. & O'Hara, M. (2012), "The Volume Clock: Insights into the High-Frequency Paradigm", *The Journal of Portfolio Management*, 39(1): 19-29.
- Funk, R. J. & Hirschman, D. (2014), "Derivatives and Deregulation: Financial Innovation and the Demise of Glass-Steagall", *Administrative Science Quarterly*, 59(4): 669-704.
- Fama, E. F. (1991), "Efficient Capital Markets: II", *The Journal of Finance*, 46(5): 1575-1617.
- Foster, G. (1979), "Briloff and the Capital Market", *Journal of Accounting Research*, 262-74.
- Gilbert, E., & Karahalios, K. (2010, May), "Widespread Worry and the Stock Market", in The International AAAI Conference on Web and Social Media (ICWSM):. 59-65), <http://www.icwsm.org/2010/papers.shtml> accessed on February 24, 2018.
- Jin, X., Shen, D. & Zhang, W. (2016), "Has Microblogging Changed Stock Market Behavior? Evidence from China", *Physica A: Statistical Mechanics and its Applications*, 452: 151-56.
- Moat, H. S., Preis, T., Olivola, C. Y., Liu, C. & Chater, N. (2014), "Using big Data to Predict Collective Behavior in the Real World", *Behavioral and Brain Sciences*, 37(1): 92-93.
- Nagata, S. & Inui, K. (2014), "Does High-speed Trading Enhance Market Efficiency? Empirical Analysis on "Arrowhead" of the Tokyo Stock Exchange", *The Journal of Trading*, 9(4): 37-47.
- Snijders, C., Matzat, U. & Reips, U. D. (2012), "Big Data: Big Gaps of Knowledge in the Field of Internet Science", *International Journal of Internet Science*, 7(1): 1-5.
- Seth, T., & Chaudhary, V. (2015), Big Data: Algorithms, Analytics, and Applications. Sound Parkway, Boca Raton, FL, CRC Press
- Seddon, J. Jonathan. & Currie, W. L. (2017), "A Model for Unpacking Big Data Analytics

- in High-frequency Trading”, *Journal of Business Research*, 70: 300-07.
- Sharpe, W. F. (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”, *The Journal of Finance*, 19(3): 425-42.
- Seddon, J. J. & Currie, W. L. (2017), “A Model for Unpacking Big Data Analytics in High-Frequency Trading”, *Journal of Business Research*, 70: 300-07.
- Singh, A. P. (2016), “R&D Spillovers & Productivity Growth: Evidence from Indian Manufacturing”, *The Indian Journal of Industrial Relations*, 51(4): 563-79.
- Singh, A. P. (2016), “Do Technology Spillovers Accelerate Performance of Firms? Unravelling a Puzzle from Indian Manufacturing Industry”, *Annals of the University Dunarea de Jos of Galati: Fascicle: XVII, Medicine*, 22(3).
- Shorter, G. W. & Miller, R. S. (2014), High-frequency trading: background, concerns, and regulatory developments (Vol. 29). Washington, DC: Congressional Research Service.
- Tian, X., Han, R., Wang, L., Lu, G. & Zhan, J. (2015), “Latency Critical Big Data Computing in Finance”, *The Journal of Finance and Data Science*, 1(1): 33-41.
- Ye, M. & Li, G. (2017), “Internet Big Data and Capital Markets: a Literature Review”, *Financial Innovation*, 3(1): 6.
- Zhang, X., Fuehres, H. & Gloor, P. A. (2011), “Predicting Stock Market Indicators Through Twitter ‘I hope it is not as bad as I fear’”, *Procedia-Social and Behavioral Sciences*, 26: 55-62.
- Zhang, W. & Skiena, S. (2010), “Trading Strategies to Exploit Blog and News Sentiment. In The International AAAI Conference on Web and Social Media (ICWSM): 2010, <http://www.icwsm.org/2010/papers.shtml> accessed on February 24, 2018