



Research in the Application of Big Data in the Field of Stock Investment

Wanying Qi¹(✉) and Michelle Sun²

¹ NAIWAI KING'S College School, Wuxi 214000, China
qiwanying1@gmail.com

² University of California Irvine, Irvine CA 92697, USA
miches6@uci.edu

Abstract. In some fields like stock investment, every decision matters. To decrease the risk that may be caused by monitoring large quantities of information artificially and increase the security, investors usually use Big Data to process information. In this paper, we are going to discuss how Big Data works in stock investment, the model and theory that is used in the application of Big Data investment, how the prospect of Big Data Investment is, what the advantages and the disadvantages of using Big Data in stock investment are, and how Big Data makes the stock investment more efficient. The data and information indicate that the diversity, accuracy and high-speed of Big Data makes stock investment more efficient. In investment, Big Data works in 3 dimensions which are momentum, value and profitability, by using different models like Behavioral Event Analysis, Funnel Analysis and Retention Analysis Model. Although it is acknowledged that Big Data is quite useful and beneficial in stock investment, it can not be denied that the use of Big Data still needs some improvement.

Keywords: Big Data · investment · finance · Artificial Intelligence

1 Introduction

In the information age, it has become impossible for humans to monitor all the information. Fortunately, with the development of artificial technology, Big Data has become a useful technology in many fields especially in Finance. Many financial institutions have successfully use the big data in financial models, such as machine learning, text sensitivity analysis to understand customer needs and preferences. There is a lot of previous literature focusing on the risk management and fraud detection. These literature think that building a fraud risk monitoring and management system mainly based on real-time big data processing is an effective way to eliminate the asymmetric information between traders. Big Data use a variety of models and formulae to conduct financial analysis, which definitely increases the efficiency and safety of areas like risk investment. But few literature study the data privacy and the regulatory compliance, which are the problems the public today are worrying about. Therefore, too many advantages of big data have been discussed. But the defects of the big data also count in this era of big data.

In this paper, we are going to talk about the application of Big Data in stock investment. We will discuss how Big Data works in stock investment, what the prospect of Big Data Investment is, what the advantages and the disadvantages of using Big Data in stock investment are, and how Big Data makes the stock investment more efficient. We get most the information online, we read different papers. This paper will show the advantages and disadvantages of using Big Data in stock investment, and help people to improve the shortcomings of Big Data. And at the end of this paper, we will discuss the future and our prospect of the big data technology.

2 Mathematical Modeling

In order to interpret the application of big data in stock investment, this logic of capital assets' pricing should be clarified at first. The most classical model of capital assets' pricing is CAPM (Capital Asset Pricing Model):

$$E(R_i) = R_f + \beta_i(E(R_m) - R_f) \tag{1}$$

where $E(R_i)$ means the expected return on the capital asset, while $E(R_m)$ is the expected return of the market. These two parameters should be estimated by the historical data. R_f is the risk-free rate of return, seen as the given parameter (government bonds).

Next, β_i is the sensitivity of the expected excess asset returns to the expected excess market returns:

$$\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} = \rho_{i,m} \frac{\sigma_i}{\sigma_m} \tag{2}$$

Thus, combining (1) and (2), the following formula could be derived:

$$E(R_i) - R_f = \rho_{i,m} \frac{\sigma_i}{\sigma_m} (E(R_m) - R_f) \tag{3}$$

Meanwhile, $E(R_i)$ or $E(R_i) - R_f$ is the dependent variable expected by the investors. $\rho_{i,m} \frac{\sigma_i}{\sigma_m}$ can be seen as a given parameter, calculated by the historical data. Thus, $E(R_m) - R_f$ or $E(R_m)$ is the only variable that needs to be estimated for an accurate $E(R_i)$.

Therefore, the difference among the expected rates of return on a certain capital asset from various investors reflects the difference among the expected rates of return on the market from those investors. In other words, the accuracy of expected rate of return on a capital asset only concerns the accuracy of the expected rate of return on the market. Thus, the expected rate of return on the market could also be seen as the sum of the actual rate of return on the market and errors:

$$E(R_m) = R_m + \varepsilon_i \tag{4}$$

where ε_i is the errors that investor i make an estimate over the rate of return on the market. R_m is the actual rate of return on the market. In a theoretical market, there are countless traders who have own independent and identical estimates over the rate of return on the

market. Thus, their errors are also independently and identically distributed while its variance σ is also larger than 0:

$$\lim_{n \rightarrow \infty} P[\sqrt{n}(\bar{\varepsilon}_n - \mu) \leq z] = \lim_{n \rightarrow \infty} P\left[\frac{\sqrt{n}(\bar{\varepsilon}_n - \mu)}{\sigma} \leq \frac{z}{\sigma}\right] = \Phi\left(\frac{z}{\sigma}\right). \quad (5)$$

Thus, as n approaches infinity, the random variables $\sqrt{n}(\bar{\varepsilon}_n - \mu)$ converge in distribution to a normal $N(0, \sigma^2)$:

$$\sqrt{n}(\bar{\varepsilon}_n - \mu) \sim N(0, \sigma^2) \quad (6)$$

In other words, as the number of the identical and independent random variable ε increases, the mean of normalized ε would approach to zero. Therefore, the estimate of rate of return on the market would be close to the real level.

The application of big data would help and accelerate this work. Taking mass data from the whole internet, quantitative practitioners would relate various factor to the rate of return on the capital assets by applicable statistical models. Enlarging database for estimation, simulation techniques would help them exam multiple reactions of the market in one time, given a certain event. With the mentioned assistance, the technique of big data could simulate the theoretical environment of Central Limit Theory, leading the investor to the intrinsic value of capital assets as possible.

3 Realistic Influence

The statistics of exact number for financial institutes that employs Big Data Technology is difficult though, Hedge Funds can be seen as a good proxy for this analysis. Because hedge funds normally employ quantitative techniques to exploit the arbitrage opportunity at which big data is a necessity.

As shown in Fig. 1, the AUM (Assets Under Management) and market size of hedge funds have risen significantly during 2011 to 2020. Meanwhile, AUM of hedge funds increased from 1,408.39 billion to 3,826.3 billion with an annual growth rate of 10.51%.

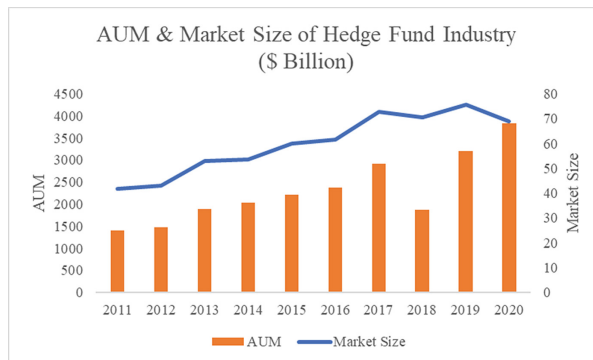


Fig. 1. AUM & Market size of Hege Fund Industry (\$ billion) Data source: Barclays Investment.

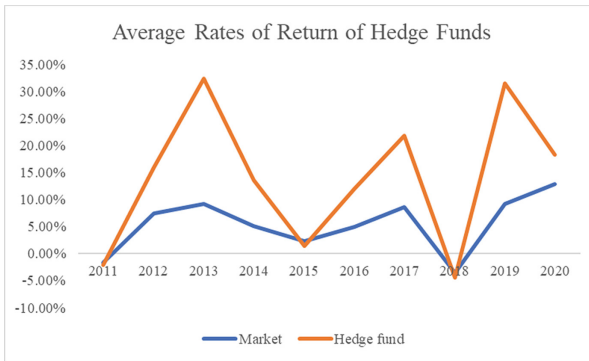


Fig. 2. Average rates of return of Hege Funds Data source: Barclay Hedge Indices

In contrast, their market size grew from 41.63 billion to 68.95 billion with a growth rate of 5.18% in the same period. The management fee of hedge funds would influence its market size directly, combining with AUM. Compared to the growth rate of AUM, that of market size in hedge funds is smaller for the past 10 years. It means that the wide application of big data in investment industry made its costs decline continuously. The rapid growth of hedge funds could reflect that big data technology has been more and more widely accepted by the investors.

In aspect of rate of return, hedge funds also outperformed during the past 10 years, compared to the market (S&P 500 Index as a proxy). As shown in Fig. 2, hedge funds that widely adopt big data technology realized an annual average rate of return of 14.07% in the past 10 years. On the contrary, that of the whole market is only 5.48%. Moreover, the hedge funds would outperform during the bull market while underperform in the bear market. Because of the application of big data, investors who employ quantitative analysis would react to the change in the market more swiftly. As a result, they could utilize better the gap between the price and the intrinsic value of the underlying assets, if making a correct expectation. If making a wrong decision, they might also take a larger loss, compared to the investors who employ traditional investment analysis methods.

In a summary, Big Data is more and more widely applied in the field of stock investment with a rapid growth. Due to its generalization, its costs also decline continuously in the past year. Its application would assist investors maximize profits by difference between price and value of a given asset with a potential probability that brings them a larger loss.

4 The Application and Prospect of Big Data in Stock Investment

4.1 How Big Data Makes Investment More Efficient

As we all know, one of the most important reasons for an inefficient financial market is the existence of asymmetric information. The situation of the market has become more and more difficult to predict in traditional ways, which makes it more difficult to make investment decisions for enterprises. With the development of the global economy, even

a small investment mistake will bring big losses to the company [1]. But the technique of the big data can help solve this problem. It is acknowledged that Big Data makes investment more efficient. But, how does it make it? With the help of big data, stock investment decisions can be divided into three parts. First, many investment institutions try to establish the database of investment decision-making platform to update the real-time data. And then, they will use the method of data mining to develop a decision model, which can promote the combination of database and model for decision. As for the final step, the decision maker will take the stock project as an example into consideration. Mathematically, this is called data simulation, which can realize investment decision. And the rich social connecting network today has increased people's social connecting efficiency by 10 times than before [2]. There are billions of messages in platforms like Facebook and Twitter. Some funds companies utilize the text sentiment analysis to subtract the investment propensity of public. And the empirical study shows that many financial firms can get a higher return in this method to buy or sell many stocks in stock markets. As for models, there are many types, like Behavioral Event Analysis, Funnel Analysis and Retention Analysis Model. In addition, using Big Data, investors can realize the centralization of information collection, information transmission and information processing [3]. Investors are allowed to find the least risky stock and make many quick decisions to avoid fund loss. For example, companies like Lear Capital use Big Data to provide any information about precious metal, acquaint users with how the market works, and allow investors to invest in the precious metal industry, which has a low risk. In some investment organizations, Big Data is used to evaluate the credit risk assessment. This allows the investors to avoid unpredictable fund loss. In this case, the risk of investment can be reduced and the efficiency can be enormously increase. In conclusion, since the high precision of the Big Data, it allows investors to analyze more accurately and make investment more efficient. What's more, helping people to analyze and suggesting the securer stocks allow those who are not professional to make more efficient investment.

4.2 How Big Data Works in Stock Investment

In financial markets, we often call the "bid data method" as the quantitative investment strategies approach. So the big data usually works in 3 dimensions: momentum, value and profitability.

In the first dimension, we often use the Fama-French three-factor model to calculate the excess return of a stock. Based on sentiment analysis, stock trends and relationships among economic factors, a more advanced machine learning techniques with big data can better recognize the relationships between companies. Leippold use the machine learning method to find the difference of Chinese and US stock market, and he finds that the biggest difference between the Chinese market and that of US is the high predictability of large or state-owned stocks over longer horizons [4].

In the second dimension, the big data can be used to determine its intrinsic value by analyzing a lot of industry-specific data that you can not find in a company's normal financial statements. Sun has collected the operating income data of a listed company in recent 10 years and he simulated these data to verify the model efficiency in processing

dynamic data and the stability of investment income [5]. Without accurate cash flows in future, we cannot calculate the net present value of a firm.

In the third dimension, as we mentioned before, we can utilize the big data to evaluate the company's web traffic patterns, which can identify which business is gaining e-commerce market share in real-time. In a word, the big data can be used in stock investment in three dimensions.

4.3 The Advantages and Disadvantages of Using Big Data in the Stock Investment

The big data can relieve the asymmetric information problem in the stock market. As is known to us all, these listed firms know more information about themselves than the investors. With big data technology, it's more easy for investors to acquire the stock transaction data than before. So it is obvious that Big Data is quite beneficial in the stock investment, however, every coin has two sides, it also has many defects. Artificial analysis and calculation always have the disadvantage of low speed in comparison to AI, the application of Big Data can assuredly increase the speed of information analysis. In 2014, Kensho financed 15 million dollars from Gaosheng Investment Bank, and use it to construct platforms for Big Data analysis and storage system [6]. Big Data has the feature of covering a wide area and the scale of data is enormous. According to IDC report, in 2018, the volume of global data is roughly 33ZB, it is predicted that the volume can increase to 175ZB by 2025. Another advantage of Big Data is that it has a wide range of varieties, high truth and availability [3]. However, the challenges of cultural variation, integrating legacy systems and a lot of expenses related to staffing, hardware, maintenance are the main disadvantages we encountered [7]. Moreover, it is hard for computers to understand the feelings of investors accurately. Human's emotions are difficult to be converted into computer language, analysts find trouble in analyzing unstructured information like investors' emotions, different language habits and geographic positions. Another disadvantage is that Big Data focus on the correlation instead of the causality. This defect may cause some information useless and made the investment less efficient. What's more, it is not feasible to ignore the human factors in the individual share [7]. There's still some concern from users of Big Data Investment APP that Big Data may invade their privacy. Other distinct disadvantages of big data are cyber security risks and the difficulties of law regulations. Many financial regulations are implementing policies to monitor the behaviors of financial institutions using the big data. When people browse the websites, it's very common to provide us personal information to sign in and enjoy more service. This practice of gathering personal data is becoming more and more widespread. However, the rapid developments of this practice have gone beyond the the current privacy regulations in big data, which leaves space for grey areas and uncertainties that cannot be solved by referring to the law. And the public become more cautious to provide their individual private information. In a word, Big Data Investment helps investors to make more efficient investment, however, there is still much room for Big Data Investment to improve. But according to the No Arbitrage Theory in financial markets, if everyone uses this big data method to make excess profits, the effectiveness of this method will be questioned.

4.4 The Prospect of Big Data Application in Investment

Many countries have successfully formulated and implemented strategic documents of big data to vigorously promote the development and application of big data. And I think there exist three trends of big data.

First, the international big data industry will be in a process of transformation. Internet, finance, telecommunication, medical treatment and government are the key areas of big data operation. However, the development and application of big data in most fields are still in the primary stage [8]. So there must have many big data firms to do this specialized work- focus on the data of industries. There exist many traditional big data companies now like Bloomberg, Wind, etc. And it's easy to form a monopolistic power in the data market since the more unique data a firm have, the cost to get it will be very high in the market. And it will increase the threshold of the big data. In the field of outsourcing, Big Data technology has become the "next big event" in the information technology industry. Currently, some outsourcing giants have begun to enter the big data market to share this big cake.

Second, the Visualization will promote the popularization of big data. The big data has a big volume. So it must have much irrelevant information. In investment, visualization can tell you directly the stock's historical returns and the prediction based on the big data technology. To utilize these useful information, firms should visualize these data. In fact, visualization can greatly narrow the distance between big data and ordinary people. It is more plausible for people to understand something in graphs than just figures. Even for people who don't know about the techniques can better understand the effect and value of big data and its analysis, so that they can fully meet the value of large data both from the national economy and from people's livelihoods.

Third, Big data security and privacy are worrying. Privacy means a lot to people today. Many studies shows that people are paying more attention to their personal information. Many Internet firms utilize their customer information to customize a specialized interface in their App. For users. But many users consider it disgusting and damage their private information. So in the near future, the information regulation must be more severe. In the research and development of large scale data, there must be a fundamental proportion to the corresponding security research, and the propulsion force for safety improvement is a negative study on the attack and abuse of big data [9]. The future investment based on the big data will focus on more aspects. Investment specialist often has his/her own way to analyze a firm. But this method is always limited to the specialist's knowledge of a specific subject. With big data, investment is more like a interdisciplinary. Many other information we tend to neglect will be captured for investors to make a wise investment decision.

5 Conclusion

These information indicates that Big Data has a number of advantages compared to humans, as the development of AI, Big Data can complete many tasks or make up some models that are complex or even impossible for humans to accomplish. However, it can not be denied that Big Data is still not perfect, which means that there's some disadvantages of using Big Data in stock investment. Even though, Big Data still has

great advantages in stock investment in the future. Many firms and scientists devote to the development of it and more and more people are now supporting the application. This paper focuses on the application of Big Data in the stock investment, however, there's still some limitation. We did not have any research in the other kind of risk investment, and all the information is from former papers. It's interesting to study the relationships between the big data and the decision making instead of only the investment decision making, so in the future, we decide to do further research in this field, we would like to use more methods to collect information.

Acknowledgments. This research is supported by Ms Alisa Wang. Thanks for the help of Ms Wang.

References

1. Pan L., Yi S.P. A study on supply chain investment decision-making and coordination in the big data environment *Ann. Oper. Res.* 2017(6), pp. 1–19
2. Kapoor, K.K., Tamilmani, K., Rana, N.P. et al. Advances in Social Media Research: Past, Present and Future. *Inf Syst Front* 20, 2018: 531–558.
3. Liu Yan. Analysis on the application of mobile big data in the financial field based on the background of big data. *Business News*.21, 2021: 75–77.
4. Markus Leippold, Qian Wang, Wenyu Zhou, Machine learning in the Chinese stock market, *Journal of Financial Economics*, 2021.
5. Chenghao Sun, Research on investment decision-making model from the perspective of “Internet of Things + Big data”, *Future Generation Computer Systems*, Volume 107, 2020, pp286–292, ISSN.0167–739X, <https://doi.org/10.1016/j.future.2020.02.003>.
6. Lv Jianlin. Application of stock investment based on big data technology. *Productivity Research*. 2019 (06): 54–57+67. doi: <https://doi.org/10.19374/j.cnki.14-1145/F.2019.06.007>. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2019&filename=SCLY201906007&v=wJffwzczpD1nqFb0KGeYaNe6GbRgKgLSueTptm6t%25md2Bac2gY7WYyT2H5u41wZvg8SI>
7. Kristoffer Bonheur, Advantages and Disadvantages of Big Data, 2019:03–21. <https://www.profolus.com/topics/advantages-and-disadvantages-of-big-data/>
8. Chang, V. (2015a). Towards a Big Data system disaster recovery in a Private Cloud. *Ad Hoc Networks*, 2015: 35, 65–82.
9. Hirsch, D. D. (2013). Glass house effect: big data, the new oil, and the power of analogy. *The Maine Law Review*, 2013: 66, 373.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

