Journal of Economic Theory and Business Management
Journal Home: https://jetbm.suaspress.org/ | CODEN: JETBAU
Vol. 2, No. 2, 2025 | ISSN 3006-4953 (Print) | ISSN 3006-4961 (Online)

Big Data-Driven ESG Quantitative Investment Strategy

WANG, Bingxing [1*]

[1] Shanghai Jingzhuo Investment Management Co., Ltd., China

**** WANG, Bingxing is the corresponding author, E-mail: stellawang6262@foxmail.com***

Abstract: With sustainability becoming more important worldwide, investors are looking more closely at environmental, social, and governance (ESG) factors. This paper looks at how big data could help investors use ESG information effectively in quantitative investing. It discusses how the use of big data techniques can lead to more accurate and transparent ESG analyses. Using regression models, the study identifies a positive relationship between companies' ESG scores and their expected stock returns. It also illustrates how detailed big data analysis can enrich the evaluation of corporate ESG performance. Despite these advantages, the practical use of such methods still faces several significant hurdles. Data quality issues, a lack of standardized ESG metrics, and dynamic market conditions can undermine model accuracy and stability. To address these obstacles, we propose improved data cleaning procedures, the promotion of industry-wide ESG standards, and enhancements to model adaptability. Looking forward, new technologies such as artificial intelligence and blockchain are likely to help ESG investing become simpler and more efficient. Using these tools can make processing ESG data faster and clearer, giving investors stronger support when making decisions. In general, applying big data in ESG investing can help promote sustainability in financial markets and may also offer steady, long-term returns. Yet, there's still uncertainty about how easily these technologies can actually be used in practice.

Keywords: Big Data, ESG Investment, Quantitative Investment Strategy, Regression Analysis, Machine Learning, Data Analysis, Sustainable Investment.

Disciplines: Finance. **Subjects:** Investment Banking.

DOI: https://doi.org/10.70393/6a6574626d.323837 **ARK:** https://n2t.net/ark:/40704/JETBM.v2n2a02

1 INTRODUCTION

In today's global financial environment, ESG considerations have become an integral part of investment decision-making [1]. Driven by the emphasis on sustainable development, companies are increasingly prioritizing environmental and social responsibilities in their operations. This shift has led investors to pay closer attention to ESG factors. Meanwhile, the rapid development of big data technology has made it feasible to apply quantitative approaches to ESG investing. By conducting precise data analysis [2], investors can now more effectively identify potential investment opportunities and risks.

Recent studies show a positive correlation between companies' ESG performance and their financial returns. For example, Friede et al. (2015) found that strong ESG performance is often associated with lower capital costs and enhanced corporate reputation [3]. These findings suggest that quantitative investment strategies supported by big data analytics can systematically incorporate ESG factors to better predict and optimize returns [4]. Such strategies involve not only an in-depth examination of traditional financial metrics but also the integration of diverse non-financial data sources—such as social media sentiment and environmental impact indicators—into the investment analysis.

In the following, we first provide an overview of ESG and explain its importance in the context of big data (Section 2). We then discuss the application of big data techniques in ESG analysis (Section 3) and introduce a quantitative ESG investment model, examining its current use (Section 4). After that, we outline the practical challenges of implementing big data-driven ESG strategies (Section 5). Finally, we conclude with insights and recommendations for future research (Section 6).

2 ESG OVERVIEW AND BACKGROUND

ESG Definition and Importance. ESG factors have become critical for evaluating a corporation's sustainability and long-term value in the modern investment landscape. Understanding what ESG entails and why it matters is fundamental to building quantitative investment strategies. ESG refers to the evaluation of a company's environmental practices, social responsibilities, and governance structure to gauge its sustainability potential in the market. Research shows that companies with high ESG ratings often outperform their peers financially. Specifically, firms with

strong ESG performance tend to exhibit lower stock volatility and achieve higher long-term returns, providing a solid theoretical basis for including ESG factors in investment portfolios. As shown in **Figure 1**, strong performance across the environmental, social, and governance dimensions is linked to greater long-term corporate value.

In this context, the advent of big data technology offers new perspectives and methods for ESG analysis. Big data is characterized by massive volume, variety, and velocity, enabling investors to obtain rich information about a company's ESG performance from diverse sources (e.g., social media, corporate reports, news feeds). However, the multidimensional and unstructured nature of these data requires advanced analytical capabilities. By using tools such as machine learning (ML) and natural language processing (NLP), analysts can significantly improve the efficiency and accuracy of ESG data processing. These techniques help in constructing and optimizing quantitative ESG investment models with far greater precision than traditional approaches.

Despite these advancements, converting ESG data into effective investment strategies remains challenging [7]. Even as data collection and analysis capabilities improve, quantitative models can be undermined by issues of data quality and bias [8]. For example, inconsistent or inaccurate ESG ratings from different providers may introduce noise, leading to erroneous conclusions. Moreover, dynamic market conditions can quickly reduce a model's applicability, requiring investors to constantly monitor and adjust their strategies to maintain effectiveness.

In summary, big data-driven ESG investment strategies are rapidly evolving. They combine robust theoretical frameworks with cutting-edge technologies to provide more scientific support for investment decisions [9]. Moving forward, researchers and practitioners must place greater emphasis on improving data accuracy and enhancing model adaptability. Such focus will help ensure that ESG quantitative strategies remain effective and reliable in an increasingly complex financial market environment. **Table 1** provides an overview of the analytical framework, highlighting the key aspects of ESG integration, big data technology, challenges, and future research focus in quantitative ESG investing.

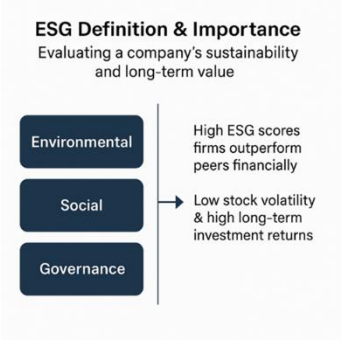

FIGURE 1. THE RELATIONSHIP BETWEEN ESG DIMENSIONS AND CORPORATE LONG-TERM VALUE

TABLE 1. ANALYTICAL FRAMEWORK FOR BIG DATA-DRIVEN ESG QUANTITATIVE INVESTMENT STRATEGY

Aspect	Key Points
ESG Definition and Importance	• Evaluating corporate sustainability and long-term value • High ESG scores correlate with better financial performance • Lower stock volatility and higher long-term returns
Introduction of Big Data Technology	• Provides new perspectives and methods for ESG analysis • Improves data processing efficiency and accuracy
Challenges in ESG Investment Strategy	• Difficulty translating ESG data into effective strategies • Data quality and bias issues can distort models
Future Research Focus	• Market dynamics can reduce model applicability • Need to improve data accuracy and model adaptability

3 APPLICATION OF BIG DATA IN ESG ANALYSIS

Big data technology is revolutionizing the way investors analyze ESG factors. The availability of large and diverse datasets, coupled with advanced analytics, provides unprecedented opportunities for evaluating corporate sustainability with greater precision [10][11]. Investors can integrate information from a wide range of sources—such as social media sentiment, real-time news feeds, supply chain records, and environmental monitoring data—to inform their ESG investment decisions. This wealth of data strengthens the foundation for constructing ESG-focused investment strategies. For instance, **Figure 2** illustrates a framework for ESG analysis driven by big data.

Natural language processing (NLP) and machine learning (ML) techniques can be applied to extract meaningful ESG insights from unstructured data. By analyzing textual content like annual reports, environmental disclosures, and news articles, these tools quantify a company's performance in environmental, social, and governance dimensions [12]. Research has shown that sentiment analysis can be used to derive an "ESG sentiment score" for a firm over time, providing investors with a valuable indicator for decision-making.

Big data methods also enable the dynamic tracking of companies' ESG performance. For example, investors can employ data mining and regression analysis on historical ESG and financial datasets to assess how past ESG performance correlates with stock returns. Such analysis allows the development of predictive models, moving ESG evaluation beyond static ratings toward a more dynamic, trend-based approach.

At the same time, analysts must consider data quality and privacy issues when leveraging big data for ESG analysis. Inaccurate or biased information can mislead investors and skew risk–return assessments. Therefore, rigorous data cleaning and validation are required before feeding data into models, to ensure the results are reliable. Additionally, it is critical to comply with data privacy regulations and maintain transparency in how ESG data are collected and used.

Looking ahead, ESG quantitative strategies are expected to become increasingly intelligent and automated [13]. As emerging technologies like AI and blockchain continue to advance, the efficiency of ESG data acquisition and processing will further improve. Greater automation and smarter analytics will enhance transparency and build trust in the ESG evaluation process. In complex market environments, these big data-driven tools will enable more accurate trend forecasts and more informed investment decisions.

Overall, incorporating big data into ESG analysis broadens the scope and depth of information available to investors while providing a stronger theoretical and empirical basis for ESG investment strategy design [14]. As technology progresses, ESG quantitative strategies will likely become even more scientific, systematic, and forward-looking, thereby making greater contributions to sustainable development goals.

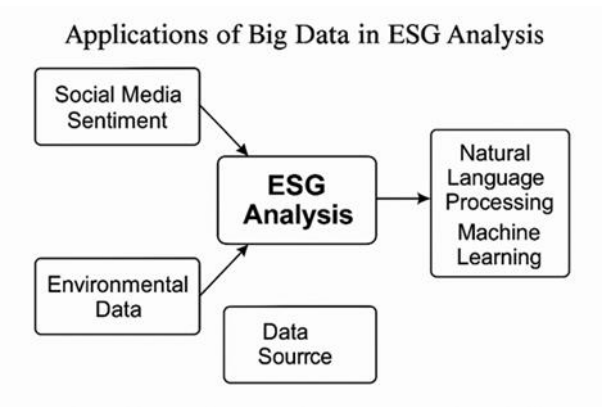

FIGURE 2. BIG DATA-DRIVEN ESG ANALYSIS FRAMEWORK

4 ESG QUANTITATIVE INVESTMENT STRATEGY

In recent years, investors have increasingly incorporated ESG metrics into quantitative investment strategies. Many investment management firms now develop quantitative models that include ESG performance to identify companies with strong social responsibility and sustainable growth potential. Constructing an ESG-driven quantitative strategy typically involves several key steps: data acquisition and processing, factor construction (including ESG-related factors), and regression analysis for return prediction. A critical component in this process is determining a reliable ESG score for each company. Practitioners often combine ESG scores with traditional financial indicators to build a multi-dimensional evaluation framework, ensuring that ESG information is effectively utilized in stock selection.

Regression Analysis Model. To quantitatively assess the impact of ESG factors on stock returns, a common approach is to use a regression model that relates expected stock returns to ESG scores (along with other variables). For example, a simple linear model can be specified as:

$$E(R\_i) = \alpha + \beta \times ESG\_i + \varepsilon\_i,$$

where $E(R\_i)$ is the expected return of stock i, ESG_i is the ESG score of stock i, α is the intercept (constant term) capturing other factors affecting returns, β is the coefficient indicating how strongly the ESG score influences returns, and ε is the error term representing variation not explained by the model. In this framework, a positive β would suggest that higher ESG scores are associated with higher expected returns. For instance, a company with an ESG score of 80 might be predicted to have a significantly higher expected return than one with a lower ESG score, providing theoretical justification for favoring high-ESG firms in a portfolio [15]. **Figure 3** illustrates the concept of an ESG quantitative investment model.

To illustrate with an example, **Table 2** presents hypothetical regression results for three companies (A, B, and C), including their ESG scores and the corresponding model parameters. Even though these values are for demonstration purposes, they reflect how an ESG factor can be quantitatively related to returns. In practice, investors would estimate β using historical data across many companies or time periods, rather than for a single firm in isolation.

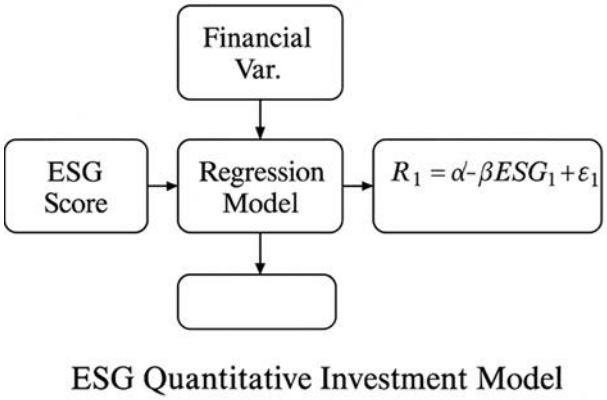

FIGURE 3. ESG QUANTITATIVE INVESTMENT STRATEGY MODEL

TABLE 2. EXAMPLE REGRESSION MODEL RELATING ESG SCORES TO EXPECTED STOCK RETURNS

Stock	ESG Score	Coefficient (β)	Intercept (α)	Error (ε)
A	80	0.15	0.05	0.02
B	75	0.10	0.07	0.01
C	85	0.12	0.03	0.03

Through regression analysis like this, investors can quantify the relationship between ESG attributes and

financial performance. The insights from such models provide a more scientific basis for investment decisions, helping investors adjust their portfolios to emphasize companies with favorable ESG profiles.

5 CHALLENGES OF ESG QUANTITATIVE INVESTMENT STRATEGIES

Despite the great potential of big data-driven ESG investing, several challenges limit their effectiveness in practice:

Data Quality. ESG datasets are often compiled from diverse sources and can include subjective or inconsistent information. For instance, one company's ESG profile might be assessed differently by various rating agencies, leading to divergent ESG scores. Poor data quality introduces noise and bias into models, which can distort results and lead to misguided investment decisions. Ensuring the accuracy, consistency, and reliability of ESG data remains a major challenge.

Lack of Standardization. There is currently no universally accepted standard for measuring or reporting ESG performance. Different rating agencies and data providers use distinct frameworks and weighting schemes — one may emphasize corporate governance while another prioritizes environmental factors. These discrepancies make it difficult to compare ESG scores across sources and to integrate them into a single model. To improve quantitative ESG strategies, the investment industry needs to move toward unified ESG evaluation standards, which would enhance comparability and confidence in the data.

Market Acceptance. Even as ESG investing gains popularity, some investors (particularly traditional ones) remain skeptical about its value. ESG-based strategies are sometimes perceived as potentially sacrificing short-term financial gains for social or ethical goals. This skepticism can lead to lower market acceptance of ESG quantitative strategies, limiting the flow of capital into ESG-aligned investments. Overcoming this challenge requires demonstrating that incorporating ESG factors can enhance risk-adjusted returns and is not merely a moral trade-off.

Dynamic Market Conditions. The relationships uncovered by ESG-focused quantitative models are not static; they can change as market conditions evolve. Unexpected events — such as regulatory shifts, economic crises, or environmental disasters — can alter a company's ESG performance or investors' perceptions of ESG importance, thereby impacting stock returns. As a result, models must be continually updated and validated against new data. Investors need to regularly recalibrate their strategies to maintain effectiveness under changing conditions, ensuring that ESG signals remain relevant in different market environments.

6 CONCLUSION

This study examined how big data technologies can enhance ESG-oriented quantitative investment strategies. We found that integrating ESG factors into quantitative models—supported by robust data analytics—can clarify the link between a company's sustainability performance and its financial returns. Consistent with prior research, companies with strong ESG performance tend to benefit from lower capital costs, better reputations, and potentially higher returns, providing a compelling rationale for including ESG metrics in investment decision-making.

Throughout the paper, we highlighted the evolving landscape of ESG investment analysis. We illustrated how big data tools (such as machine learning and NLP) allow investors to extract insights from a vast array of structured and unstructured data sources, thereby improving the accuracy and transparency of ESG assessments. Our regression model example demonstrated in quantitative terms the positive impact that superior ESG scores can have on expected stock returns, reinforcing the value of ESG integration in portfolio management.

At the same time, we identified several practical challenges that must be addressed to fully realize the benefits of big data-driven ESG strategies. Data quality and consistency emerged as critical concerns, given that unreliable or biased information can lead to false signals. The lack of standardized ESG reporting frameworks across the industry also hinders comparability and confidence. Moreover, we discussed how dynamic market conditions mean that the relationships between ESG factors and financial performance can shift over time, necessitating adaptable models and continuous monitoring. Focusing future research on improving data accuracy and developing more resilient, adaptable quantitative models will be essential for overcoming these hurdles and ensuring the stability of ESG investment outcomes.

Looking ahead, the continued advancement of technologies like artificial intelligence, blockchain, and the Internet of Things is poised to further transform ESG investing. These innovations will enable more efficient data collection, real-time analysis, and greater transparency, thereby increasing investors' trust in ESG data and metrics. We anticipate that ESG quantitative investment strategies will become even more intelligent and automated, helping investors identify opportunities and manage risks with greater precision. Ultimately, by strengthening analytical techniques, promoting industry-wide ESG standards, and building broader market acceptance, big data-driven ESG strategies are expected to provide investors with rigorous and forward-looking decision support. Such strategies will not only seek competitive financial returns but also contribute to global sustainable development objectives, aligning the goals of investors with the broader needs of society.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

WANG, Bingxing

Shanghai Jingzhuo Investment Management Co., Ltd., China.

REFERENCES

[1] Fu, L. (2021). The power of change — ESG "deep type" investment strategy in China. Contemporary Manager, (01), 19–20.

[2] Yang, X. (2022). Research on quantitative strategies based on investor sentiment (Master's thesis, Central University of Finance and Economics). Central University of Finance and Economics. (Supervisors: Wei Xu & Li Ruiqiang)

[3] Xie, Y. (2022, December 22). Multi-factor driven private equity institutions focus on ESG investment. Financial Times, (006).

[4] ESG strategies: Fulfilling social responsibility and practicing long-term investment. (2021). Shanghai State-owned Assets, (07), 101–103.

[5] Tang, Y. (2022). Research on quantitative investment strategies based on public opinion data (Master's thesis, Shandong University). Shandong University. (Supervisor: Shi Yufeng)

[6] Yu, D., Liu, L., Wu, S., Li, K., Wang, C., Xie, J., ... & Ji, R. (2024). Machine learning optimizes the efficiency of picking and packing in automated warehouse robot systems. In 2024 International Conference on Computer Engineering, Network and Digital Communication (CENDC 2024).

[7] Li, K., Liu, L., Chen, J., Yu, D., Zhou, X., Li, M., ... & Li, Z. (2024, November). Research on reinforcement learning based warehouse robot navigation algorithm in complex warehouse layout. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 296-301). IEEE.

[8] Li, K., Chen, J., Yu, D., Dajun, T., Qiu, X., Lian, J., ... & Han, J. (2024, October). Deep reinforcement learning-based obstacle avoidance for robot movement in warehouse environments. In 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (pp. 342-348). IEEE.

[9] Li, K., Wang, J., Wu, X., Peng, X., Chang, R., Deng, X., ... & Hong, B. (2024). Optimizing automated picking systems in warehouse robots using machine learning. arXiv preprint arXiv:2408.16633.

[10] Sun, J., Zhang, S., Lian, J., Fu, L., Zhou, Z., Fan, Y., & Xu, K. (2024, December). Research on Deep Learning of Convolutional Neural Network for Action Recognition of Intelligent Terminals in the Big Data Environment and its Intelligent Software Application. In 2024 IEEE 7th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE) (pp. 996-1004). IEEE.

[11] Mao, Y., Tao, D., Zhang, S., Qi, T., & Li, K. (2025). Research and Design on Intelligent Recognition of Unordered Targets for Robots Based on Reinforcement Learning. arXiv preprint arXiv:2503.07340.

[12] Zuo, Q., Tao, D., Qi, T., Xie, J., Zhou, Z., Tian, Z., & Mingyu, Y. (2025). Industrial Internet Robot Collaboration System and Edge Computing Optimization. arXiv preprint arXiv:2504.02492.

[13] Zhang, T. (2025). Combining Blockchain and AI to Optimize the Intelligent Risk Control Mechanism in Decentralized Finance. Journal of Industrial Engineering and Applied Science, 3(2), 26-32.

[14] Zhou, Y., Shen, J., & Cheng, Y. (2025). Weak to strong generalization for large language models with multi-capabilities. In The Thirteenth International Conference on Learning Representations.

[15] Zhou, Y., Zhang, J., Chen, G., Shen, J., & Cheng, Y. (2024). Less is more: Vision representation compression for efficient video generation with large language models.