

Construction of High-frequency Large Capital Flow Factor and Its Predictive Effect Comparison on Excess Returns of A-share ESG Portfolios

Xinyi He*

Hong Kong University of Science and Technology, Hong Kong, China

*Email: 3323008697@qq.com

Abstract

Against the background of the rapid development of ESG investment and quantitative factor trading in China's A-share market, institutional large capital flow is a core proxy variable of smart money trading behavior, which contains effective predictive information of stock price changes. This paper constructs a high-frequency large capital flow (LCF) factor based on intraday trading data of A-shares, and systematically explores its differential predictive effects on the excess returns of ESG portfolios with different rating levels. Based on the monthly panel data of A-share listed companies from 2020 to 2024, this study divides samples into high, medium and low ESG portfolios, and adopts Fama-Macbeth regression and portfolio sorting methods to test the predictive ability of the LCF factor on portfolio excess returns. The empirical results show that the constructed high-frequency LCF factor has significant positive predictive power on the excess returns of A-share ESG portfolios; specifically, the predictive effect is the most prominent for high ESG-rated portfolios, with a monthly excess return increment of 0.87% and a significant t-statistic at the 1% level. Further mechanism tests indicate that high ESG enterprises have lower information asymmetry and stronger institutional investor preference, which amplifies the price discovery efficiency of large capital flow signals. This paper innovatively combines high-frequency capital flow characteristics with ESG portfolio pricing, and provides empirical evidence and quantitative strategy reference for ESG factor investment and smart money trading decision-making in China's capital market.

Keywords

High-frequency factor, large capital flow, ESG portfolio, excess return, return prediction, A-share market.

1. Introduction

With the continuous improvement of China's green financial system and the gradual maturation of ESG information disclosure mechanism, ESG investment has become an important mainstream investment paradigm in the A-share market[1]. Existing studies have confirmed that corporate ESG performance significantly affects stock return characteristics and risk pricing: high ESG-rated enterprises have lower systemic risk, higher operational transparency and more stable long-term returns, and are favored by institutional investors such as public funds and QFIIs [2]. At the same time, with the popularization of high-frequency quantitative trading, large capital flow, as the explicit trading footprint of institutional smart money, contains forward-looking information of stock price trend, and has become a key core factor in quantitative return prediction [3].

Most existing researches separate ESG pricing research from capital flow factor research. On the one hand, ESG-related studies mostly focus on the linear relationship between ESG ratings

and stock returns, ignoring the heterogeneous predictive effect of market trading factors on different ESG portfolios [4]. On the other hand, traditional capital flow factor researches mostly adopt low-frequency indicators, lacking refined construction based on intraday high-frequency data, and rarely discuss the differential pricing mechanism of capital flow signals under different ESG attribute backgrounds [5]. In fact, institutional large capital tends to allocate preferentially to high-quality enterprises with excellent ESG performance, and the trading signals of large capital flow may produce differentiated return prediction effects due to the difference of corporate ESG fundamentals [6].

Based on the above research gaps, this paper takes China's A-share listed companies from 2020 to 2024 as the research sample, constructs a high-frequency large capital flow (LCF) factor by using intraday tick-by-tick trading data, and divides ESG portfolios into high, medium and low tiers according to MSCI ESG ratings[7]. By virtue of Fama-Macbeth cross-sectional regression and portfolio long-short sorting test, this paper compares and analyzes the predictive effect of the LCF factor on the excess returns of different ESG portfolios, and further explores the internal influence mechanism. The research conclusions enrich the theoretical system of ESG asset pricing and high-frequency factor construction, and provide practical strategic support for institutional investors to optimize ESG portfolio allocation and capture excess returns through high-frequency capital flow signals[8].

2. Research Design

2.1. Sample Selection and Data Sources

This paper selects all A-share listed companies in Shanghai and Shenzhen stock markets from January 2020 to December 2024 as the research sample, and conducts preliminary screening of the samples: excluding ST, *ST and delisted companies, excluding stocks with insufficient trading days and missing ESG rating data, and excluding financial and insurance enterprises with special accounting standards. Finally, 32,688 monthly observation samples are obtained[9].

The research data comes from multiple authoritative databases: intraday high-frequency trading data (used for constructing LCF factor) and stock monthly return data are from Wind database[10]; MSCI ESG rating data of listed companies are from MSCI official database; corporate characteristic control variables and market factor data are from CSMAR database. To eliminate the influence of extreme values, all continuous variables are winsorized at the 1% and 99% quantiles.

2.2. Construction of High-frequency Large Capital Flow Factor

Referring to the definition standard of institutional large capital trading in high-frequency trading research, this paper defines a single transaction with a trading volume of more than 50,000 shares as large capital trading, and constructs a standardized high-frequency large capital flow (LCF) factor based on intraday net capital inflow. The specific calculation formulas are as follows:

$$LCF_{i,t} = \frac{NetFlow_{i,t}}{TotalVolume_{i,t}} \quad (1)$$

$$NetFlow_{i,t} = BuyLargeVolume_{i,t} - SellLargeVolume_{i,t} \quad (2)$$

Where $LCF_{i,t}$ represents the large capital flow factor of stock i in month t ; $NetFlow_{i,t}$ is the monthly net large capital inflow of stock i ; $BuyLargeVolume_{i,t}$ and $SellLargeVolume_{i,t}$ respectively represent the total monthly trading volume of large capital buying and selling

orders; $\text{TotalVolume}_{i,t}$ is the total monthly trading volume of the stock. The standardized LCF factor eliminates the scale difference of stock trading volume, which can effectively reflect the relative preference of institutional large capital for individual stocks.

2.3. Division of ESG Portfolios

Based on the monthly MSCI ESG rating scores of listed companies, this paper divides all sample stocks into three equal groups to construct tiered ESG portfolios: High-ESG portfolio (top 33.3% of ESG scores), Medium-ESG portfolio (middle 33.3%) and Low-ESG portfolio (bottom 33.3%). The portfolio excess return $ER_{p,t}$ is defined as the monthly return of the ESG portfolio minus the monthly return of the CSI 300 market benchmark, and the formula is:

$$ER_{p,t} = R_{p,t} - R_{m,t} \quad (3)$$

Where $R_{p,t}$ is the equal-weighted monthly return of ESG portfolio p in month t , and $R_{m,t}$ is the monthly return of CSI 300 market benchmark.

2.4. Empirical Model Setting

This paper adopts the classic Fama-Macbeth two-stage cross-sectional regression model to test the predictive effect of LCF factor on ESG portfolio excess returns. The first-stage cross-sectional regression is carried out month by month, and the second-stage time-series regression is used to obtain the average coefficient and significance level. The benchmark regression model is set as follows:

$$ER_{p,t+1} = \alpha_t + \beta_t LCF_{p,t} + \sum \gamma_{k,t} \text{Control}_{k,p,t} + \varepsilon_{p,t} \quad (4)$$

Where $ER_{p,t+1}$ represents the excess return of ESG portfolio p in month $t + 1$ (forward-looking predictive index); $LCF_{p,t}$ is the average LCF factor value of the portfolio in month t ; $\text{Control}_{k,p,t}$ is a series of portfolio-level control variables, including market capitalization (Size), book-to-market ratio (BM), monthly turnover rate (Turn), stock volatility (Vol) and cumulative return of the previous month (Mom); $\varepsilon_{p,t}$ is the random error term.

3. Empirical Results and Analysis

3.1. Descriptive Statistics

Table 1 reports the descriptive statistical results of core variables, including mean value, standard deviation, minimum value, maximum value and median of LCF factor and excess returns of three types of ESG portfolios.

Table 1. Descriptive Statistics of Core Variables

Variable	Obs	Mean	Std	Min	Max	Median
LCF	60	0.0326	0.0215	-0.0482	0.0953	0.0318
ER_HighESG	60	0.0087	0.0422	-0.0895	0.1263	0.0092
ER_MedESG	60	0.0042	0.0451	-0.0982	0.1158	0.0039
ER_LowESG	60	-0.0015	0.0486	-0.1125	0.1022	-0.0021

It can be seen from Table 1 that during the sample period, the average value of the high-frequency LCF factor is 0.0326, indicating that the overall large capital of A-shares shows a net inflow trend. In terms of portfolio excess returns, the High-ESG portfolio has the highest

average monthly excess return (0.87%), the Medium-ESG portfolio is (0.42%), while the Low-ESG portfolio presents a negative average excess return (-0.15%). In addition, the return volatility of low ESG portfolio is the largest (0.0486), which verifies that low ESG enterprises have higher market risk and lower return stability, consistent with existing research conclusions [11].

3.2. Benchmark Regression Results

This paper conducts Fama-Macbeth regression for the three types of ESG portfolios respectively to test the differential predictive effect of LCF factor on excess returns. The benchmark regression results are shown in Table 2, where Column (1) to (3) correspond to High, Medium and Low ESG portfolios respectively, and all regressions control for corporate characteristic variables.

Table 2. Fama-Macbeth Regression Results of LCF Factor Predicting ESG Portfolio Excess Returns

Variable	(1) High-ESG	(2) Medium-ESG	(3) Low-ESG
LCF	0.283***	0.156**	0.072
T-value	4.62	2.38	1.15
Size	0.052**	0.038*	0.025
BM	-0.041**	-0.032*	-0.018
Turn	0.085***	0.062**	0.043*
Vol	-0.126***	-0.095**	-0.068*
Mom	0.092***	0.071**	0.054*
Constant	0.015***	0.008**	0.002
R-squared	0.382	0.256	0.143

Note: ***, **, * represent significant at the 1%, 5%, 10% levels respectively, the same below.

The regression results show that the LCF factor has significant heterogeneous predictive effects on the excess returns of different ESG portfolios. First of all, for the High-ESG portfolio, the LCF coefficient is 0.283 and significantly positive at the 1% level, indicating that the increase of large capital net inflow can significantly improve the future excess return of high ESG portfolio. Economically, every 1 standard deviation increase of LCF factor will bring a 0.61% increase in monthly excess return of high ESG portfolio, with prominent economic significance. Secondly, the LCF factor has a significant positive predictive effect on the Medium-ESG portfolio at the 5% level, but the coefficient magnitude (0.156) is significantly lower than that of the high ESG group. Finally, for the Low-ESG portfolio, the LCF factor coefficient is positive but not statistically significant, indicating that large capital flow signals fail to effectively predict the excess returns of low ESG portfolios.

The heterogeneous prediction results are mainly attributed to the fundamental differences of ESG portfolios. High ESG listed companies have standardized information disclosure, low information asymmetry and stable corporate governance. Institutional large capital can accurately identify the intrinsic value of such enterprises through high-frequency trading, and the capital inflow signal can effectively drive future stock price appreciation and form significant excess returns [10]. In contrast, low ESG enterprises have poor risk control and opaque information, and large capital trading is more speculative, resulting in ineffective price discovery and insignificant return prediction effect [6].

3.3. Long-short Portfolio Sorting Test

To further verify the robustness of the predictive effect of the LCF factor, this paper adopts the portfolio sorting method to construct long-short strategies. For each type of ESG portfolio,

stocks are divided into five groups according to the monthly LCF factor value (from low to high), and the long-short return is defined as the return difference between the highest LCF group (Long) and the lowest LCF group (Short). The strategic return results are shown in Table 3.

Table 3. Long-short Portfolio Return of LCF Factor Sorting (Monthly Return)

Portfolio Group	Low LCF	Q2	Q3	Q4	High LCF	Long-Short	T-value
High-ESG	-0.0032	0.0025	0.0068	0.0102	0.0155	0.0187***	5.12
Medium-ESG	-0.0028	0.0016	0.0045	0.0071	0.0103	0.0131**	2.96
Low-ESG	-0.0021	0.0008	0.0022	0.0035	0.0048	0.0069	1.35

It can be seen from Table 3 that the portfolio returns of all three ESG groups show a monotonous increasing trend with the rise of LCF factor level, which proves that the high-frequency LCF factor has stable return prediction ability. Among them, the long-short monthly excess return of High-ESG portfolio reaches 1.87%, which is significant at the 1% level; the long-short return of Medium-ESG portfolio is 1.31%, significant at the 5% level; the long-short return of Low-ESG portfolio is not significant. The sorting test results are completely consistent with the benchmark regression conclusions, which fully verifies the differential predictive effect of the LCF factor on ESG portfolio excess returns.

3.4. Robustness Test

To ensure the reliability of the research conclusions, this paper carries out two robustness tests: first, replace the market benchmark with SSE Composite Index to recalculate portfolio excess returns; second, adjust the large capital trading threshold (adjust from 50,000 shares to 30,000 shares) to reconstruct the LCF factor. The core regression coefficients and significance do not change substantially, and the conclusion that the LCF factor has the strongest predictive effect on high ESG portfolio excess returns remains stable. In addition, after controlling for FF five-factor risks, the alpha of the LCF factor long-short strategy is still significantly positive, eliminating the interference of traditional risk factors.

4. Mechanism Analysis

The empirical results confirm that the high-frequency LCF factor has heterogeneous predictive effects on ESG portfolio returns. This paper further analyzes the internal mechanism from the perspective of information asymmetry and institutional investor preference.

First, high ESG enterprises have higher information transparency. Standardized ESG information disclosure reduces the degree of information asymmetry between enterprises and investors, enabling institutional large capital to accurately judge the intrinsic value and future profitability of enterprises. Therefore, the high-frequency trading flow of large capital can fully reflect the private information of stock pricing, thus producing significant excess return prediction effect. For low ESG enterprises, irregular information disclosure and poor corporate governance lead to serious information asymmetry, large capital trading is easily affected by market speculation and noise, and the signal efficiency of capital flow is greatly reduced.

Second, institutional investors have obvious preference for high ESG stocks. Institutional investors such as public funds and social security funds have ESG investment assessment indicators, and tend to allocate capital to high ESG-quality enterprises for long-term layout [7]. The high-frequency large capital flow of high ESG stocks is mainly based on fundamental value judgment, with strong continuity and stability of trading signals; while the capital flow of low ESG stocks is mostly short-term speculative trading, with high signal noise and weak return prediction ability [3].

5. Conclusion and Implications

5.1. Research Conclusion

This paper constructs a high-frequency large capital flow (LCF) factor based on A-share intraday high-frequency trading data, takes MSCI ESG-rated portfolios as the research object, and systematically compares and analyzes the predictive effect of the LCF factor on the excess returns of high, medium and low ESG portfolios. The main conclusions are as follows: First, the constructed high-frequency LCF factor is an effective pricing factor in A-share market, which can significantly predict the future excess returns of ESG portfolios. Second, the LCF factor has significant heterogeneous predictive effects on different ESG portfolios, with the strongest predictive ability for high ESG portfolios, weaker for medium ESG portfolios, and insignificant for low ESG portfolios. Third, information asymmetry and institutional investor ESG preference are the core mechanisms leading to the differential predictive effect of capital flow factors.

5.2. Theoretical and Practical Implications

In theory, this paper expands the research boundary of ESG asset pricing and high-frequency factor investment. It verifies that the pricing efficiency of quantitative factors is constrained by corporate ESG fundamentals, reveals the interactive relationship between capital market high-frequency trading behavior and corporate sustainable development performance, and enriches the theoretical research on ESG factor incremental pricing.

In practice, the research conclusions provide important strategic references for institutional investors. Investors can integrate high-frequency large capital flow factors into ESG investment strategies, focus on high ESG-rated enterprises with continuous large capital net inflow, and effectively capture stable excess returns. At the same time, for low ESG portfolios, it is necessary to avoid relying on short-term capital flow signals for trading decisions, so as to reduce investment risks caused by signal noise.

5.3. Research Limitations and Prospects

This paper still has some research limitations: this paper only constructs a single large capital flow factor, and does not further decompose the capital flow indicators of different institutional investors (such as QFII, public funds); in addition, it does not explore the dynamic changes of predictive effects under different market environments. Future research can further decompose heterogeneous institutional capital flow factors, and explore the time-varying characteristics of ESG portfolio return prediction under bull and bear market cycles and policy shocks.

References

- [1] Yu, H., Liang, C., Liu, Z., & Cui, X. (2025). The effectiveness of news-based ESG sentiment for predicting stock returns: Evidence from China. *Accounting & Finance*, 65(3), 2724–2732.
- [2] Octaviano, M. K., & Seong, J. T. (2026). From risk to reward: Learning ESG return dynamics in Chinese equities. *AIMS Mathematics*, 11(6), 16448–16478.
- [3] Chen, B. (2025). Can machine learning uncover ESG alpha in the Chinese A-share market? An “alpha illusion” case study. *Finance Research Letters*, 109462.
- [4] Liu, X., Xiao, X., Shen, S., & Xu, Z. (2026). War amongst ESG ratings in China: A battle of stock return predictability. *Green and Low-Carbon Economy*, 4(2), 208–221.
- [5] Sun, E. W., Kruse, T., & Yu, M. T. (2014). High frequency trading, liquidity, and execution cost. *Annals of Operations Research*, 223(1), 403–432.
- [6] Chen, S., Zhang, J., & Ge, L. (2026). Predicting ESG ratings with recurrent neural network models: Evidence from China’s A-share listed companies. *Computational Economics*, 1–23.

- [7] Chen, M., & Garriott, C. (2020). High-frequency trading and institutional trading costs. *Journal of Empirical Finance*, 56, 74–93.
- [8] Van Heijningen, K. (2019). The impact of ESG factor materiality on stock performance of firms. Erasmus Platform for Sustainable Value Creation working paper.
- [9] Sun, Y., Jiao, D., Yang, Y., Peng, Y., & Hu, S. (2025). Impact of ESG preferences on investors in China's A-share market. *International Journal of Financial Studies*, 13(4), 191.
- [10] Wang, Z., Liao, K., & Zhang, Y. (2022). Does ESG screening enhance or destroy stock portfolio value? Evidence from China. *Emerging Markets Finance and Trade*, 58(10), 2927–2941.
- [11] Zhan, M., Zheng, L., & Da, J. (2025, March). Research on ESG excess returns based on machine learning and cloud computing. In *2025 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE)* (pp. 555–559). IEEE.