International Review of Financial Analysis 96 (2024) 103775

Contents lists available at ScienceDirect

International Review of Financial Analysis

journal homepage: www.elsevier.com/locate/irfa

# Investor attention and anomalies: Evidence from the Chinese stock market

Danyan Wen [a], Zihao Zhang [a], Jing Nie [a,*], Yang Cao [b]

[a] School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Business School, University of Sussex, Brighton BN1 9SL, UK

ARTICLE INFO

*JEL classification:*
G12
G15
G41

*Keywords:*
Investor attention
Anomalies
China's stock market
PLS
Investor sentiment

ABSTRACT

This paper investigates how investor attention influences anomalies in the Chinese stock market. Utilizing data from 2011 to 2022, we propose investor attention composite indices using the partial least squares method, combining information from 11 attention proxies. By analyzing the newly proposed index, we explore the impact of investor attention on stock market anomalies. Our results demonstrate that investor attention has a positive effect on concurrent market anomalies, a relationship that remains robust even when considering factors such as the Fama-French three factors and investor sentiment. Further examination utilizing a composite index of investor attention derived from scaled principal component analysis yields similar results. Notably, our research indicates that investor attention significantly impacts anomaly returns in the subsequent month, suggesting potential forecasting capabilities.

## 1. Introduction

The concept of investor attention originates with Kahneman (1973), indicating that attention is a scarce cognitive resource and that the limited nature of attention causes an individual's information processing efficiency to decrease as the amount of information increases. Limited attention becomes a binding force implicit in the decision-making process of investors. Investors are the main body of stock markets, and their decision-making behavior is inevitably closely related to stock price movements, making investors' attention play an essential role in the stock market and drawing much attention from scholars. For instance, an increasing body of literature incorporates investor attention as a foundational aspect in their studies (Zhu, Zhang, Wu, Zheng, & Zhang, 2021; Meshcheryakov & Winters, 2022; Que, Qin, & Zhang, 2022; Zhang, Chen, Wu, & Zhu, 2022; Lan, Xie, Mi, & Zhang, 2024). And mixed evidence regarding the impact of investor attention on stock market dynamics is provided (Andrei, Friedman, & Ozel, 2023; Ballinari, Audrino, & Sigrist, 2022; Ben-Rephael, Da, & Israelsen, 2017; Da, Hua, Hung, & Peng, 2024; DellaVigna & Pollet, 2009; Han, Hirshleifer, & Walden, 2022; Hirshleifer, Lim, & Teoh, 2009; Hu, Li, Goodell, & Shen, 2021; Zhang et al., 2023). This paper contributes to the literature by examining the impact of investor attention on anomalies in the Chinese stock market.

The relationship between investor attention and stock market anomalies is intuitive and attracts the exploration of many high-quality literature. First, investors' limited attention when making investment decisions can cause them to overlook helpful details such as earnings news (DellaVigna & Pollet, 2009), leading to delayed information processing and underreaction to relevant information (Hirshleifer et al., 2009; Peng & Xiong, 2006). Therefore, greater returns can be realized for underreaction-related anomalies once investors pay attention. Second, a higher level of investor attention can cause investors to overreact to irrelevant or useless information (Barber & Odean, 2013; Meng, Li, & Xiong, 2024) and amplify mispricing (Atilgan, Bali, Demirtas, & Gunaydin, 2020; Shue & Townsend, 2021). Whether investors underreact due to low attention or overreact due to high attention, this allocation directly affects anomaly returns, which are the return differentials across stocks. Third, investor attention can function as a coordination mechanism for arbitrageurs who might otherwise be hesitant to invest due to holding costs (Abreu & Brunnermeier, 2002) and synchronicity risk (He & Manela, 2016). Consequently, coordinated arbitrage accelerates the realization of arbitrage profits across various anomalies, regardless of whether these anomalies are caused by market underreaction or speculative investing.

China, as the world's second-largest economy, also holds the second-largest stock market globally. Understanding investor behavior and anomalies in the Chinese stock market requires attention to its unique characteristics. One standout feature is the significant presence of retail

* Corresponding author at: School of Economics and Management, Nanjing University of Science and Technology, Xiaolingwei 200, Xuanwu District, Nanjing 210094, China.
*E-mail address:* niejing0203@126.com (J. Nie).

https://doi.org/10.1016/j.irfa.2024.103775
Received 28 June 2024; Received in revised form 8 November 2024; Accepted 8 November 2024
Available online 12 November 2024
1057-5219/

investors, known for their sometimes irrational investment decisions. Research by Barber and Odean (2008) shows that these investors are active in buying stocks when market attention is high, as highlighted by Karlsson, Loewenstein, and Seppi (2009) who note their focus on positive news while avoiding negative news. These findings emphasize the substantial impact of investor attention on market dynamics in China. Another defining aspect is the existence of arbitrage restrictions, as demonstrated by Shi, Wang, Xia, and Zhen (2023), indicating that pricing anomalies stem from limitations on short selling, hindering quick corrections of overvalued stocks. This situation, combined with investors' behavioral biases, further amplifies anomalies in stock pricing. Lastly, investor sentiment plays a crucial role in the Chinese market, identified as a key factor affecting anomaly returns. Given the prevalence of individual investors in China, investor sentiment significantly influences market outcomes. This interplay between investor attention and anomalies presents a challenge in comprehending their relationship in this distinct market environment.

To fulfill our research goal, our paper proceeds in the following steps. First, we investigate the contemporaneous relationship between investor attention and market anomalies in the Chinese stock market and analyze the impact of investor attention on market anomalies. To construct the anomalies, we select the range of anomalies following Qiao (2018), who tests 231 anomalies in the Chinese stock market. Our criterion for selecting anomalies is that the anomaly returns are significant at the 5 % level. To construct investor attention, we choose 11 individual attention proxies: abnormal trading volume (Barber & Odean, 2008), extreme returns (Barber & Odean, 2008), past returns (Aboody, Lehavy, & Trueman, 2010), analyst coverage (Hirshleifer & Teoh, 2003), changes in advertising expenses (Lou, 2014), mutual fund inflows and outflows (Chen, Tang, Yao, & Zhou, 2022), media coverage (Barber & Odean, 2008), web search volume index (Que & Zhang, 2021), nearness to CSCI 52-week highs, and nearness to CSCI historical highs (Li & Yu, 2012). The dimensionality reduction method of partial least squares (PLS) is applied to aggregate information about investor attention from these 11 attention proxies to obtain a composite index of investor attention applicable to the Chinese stock market.

Second, we attempt to control for the potential influence of the Fama-French three factors (Fama & French, 1993) and investor sentiment. To be specific, we incorporate the Fama-French three factors as control variables into our benchmark model. By including these factors, we seek to account for their potential explanatory role in anomalies. As anomalies represent return differentials among stocks, the inclusion of the Fama-French three factors allows us to discern the direct effect of investor attention on anomalies more intuitively. In addition, we introduce the investor sentiment variable into our analysis. Both investor sentiment and investor attention are recognized as behavioral biases exhibited by investors. Previous studies have explored the relationship between investor sentiment and stock market anomalies, demonstrating its influence (Han & Shi, 2022; Stambaugh, Yu, & Yuan, 2012). To capture any impact of investor sentiment on our research, we include it as a control variable in our benchmark model. By incorporating the Fama-French three factors and the investor sentiment variable, we aim to effectively control for their potential effects and obtain more accurate insights into the relationship between investor attention and anomalies.

Additionally, we conduct a comparative analysis to examine the differential impact of investor attention and investor sentiment on anomalies. Our empirical findings indicate that investor attention significantly affects anomalies. To capture investor attention at the firm level, we employ the web search volume index as a proxy. To isolate the investor attention factor from anomaly returns, we construct anomalies by grouping the web search volume index and anomaly variables in a bivariate manner. By comparing the disparity in anomaly returns among different attention groups, we elucidate the distinctiveness of the impact. Specifically, we treat the bivariate grouping of anomaly returns as an explanatory variable in our benchmark model. We regress this variable on investor attention and discover that investor attention possesses greater explanatory power over anomaly returns in the high-attention group compared to the low-attention group. This finding underscores the differential influence of investor attention across varying levels of attention groups. By performing this comparative analysis, we shed light on the differing effects of investor attention and investor sentiment on anomalies, further enriching our understanding of their respective contributions.

Third, we conduct an extended analysis based on the main findings of this paper. To examine the temporal relationship between investor attention and anomalies, we analyze the relationship between investor attention and future anomaly returns. We find that investor attention has a significant positive effect on short-term anomaly returns (1 month) but fails to show significant effects on more distant anomaly returns (3 and 6 months). To ensure the robustness of our findings, we employ the scaled principal component analysis (sPCA) dimensionality reduction technique. This sPCA method allows us to construct a composite index of investor attention, providing an alternative approach to examining the relationship with stock market anomalies. Encouragingly, our analysis demonstrates that the composite index of investor attention constructed using the sPCA method yields results consistent with our main analysis, reinforcing the validity and reliability of our conclusions.

This study contributes to the existing literature in two aspects. On the one hand, our paper adds to an emerging line of research on the impact of investor attention on the stock markets. Previous studies have primarily focused on the detrimental impact of investor attention on the stock risk premium (Chen et al., 2022) and the cross-sectional stock returns (Dong, Wu, Fang, Gozgor, & Yan, 2022). Da et al. (2024) have also emphasized the relationship between investor attention and stock market returns. In contrast, our paper goes beyond this existing research by investigating the effect of investor attention on stock market anomalies in the Chinese stock markets. Our analysis reveals that investor attention can exert a significant impact on anomalies in the short term, thereby enhancing our understanding of the factors influencing market anomalies. Besides, we introduce investor sentiment into the benchmark model and conduct a comparative analysis of the influence of investor sentiment versus investor attention on anomalies. In behavioral finance, investor sentiment and attention are distinct forms of investor behavioral biases. Our empirical findings demonstrate that there is a distinction between the effects of sentiment and attention on anomalies.

On the other hand, our study deals with the measurement of investor attention. Many scholars use a single indicator to measure investor attention (Aboody et al., 2010; Barber & Odean, 2008; Da, Engelberg, & Gao, 2011; Drake, Jennings, Roulstone, & Thornock, 2017; Drake, Roulstone, & Thornock, 2015; Hirshleifer, Hsu, & Li, 2013; Hirshleifer & Teoh, 2003; Lee, Ma, & Wang, 2015; Li & Yu, 2012; Lin, 2005; Lou, 2014; Szczygielski, Charteris, Bwanya, & Brzeszczyński, 2024). However, individual attention indicators have limited power in explaining stock market returns Chen et al. (2022). We thus construct composite indices of investor attention using multiple single indicators via the PLS dimensionality reduction method. We select 11 individual attention indicators that apply to the Chinese stock markets and construct a composite index of investor attention regarding to the specific anomaly returns. This composite index offers a more holistic portrayal of investor attention within the stock markets than a single indicator.

The remainder of the paper is organized as follows: Section 2 describes the study's two core variables: investor attention and market anomalies. Section 3 reports the empirical results. Section 4 presents the extended analyses, and Section 5 concludes the study.

## 2. Data and methodology

This study aims to examine how investor attention impacts market anomalies in the Chinese stock market, with a focus on A-share stocks traded on the Shanghai and Shenzhen Stock Exchanges. To maintain data accuracy and reliability, companies labeled as ST or ST* are deliberately omitted from the research sample to prevent potential bias from abnormal data. The data for this study are sourced from various public databases, including Resset, CNRDS, Wind, and CSMAR.

### *2.1. Stock market anomalies*

This section provides an overview of 23 anomalies detected in the Chinese stock market. We analyze the average time-series returns and Newey-West *t*-statistic for these anomalies across both the full sample period and the study sample period. These anomalies stem from research by Qiao (2018), who identified 41 statistically significant anomalies at a 5 % level of significance ($|t| \geq 1.96$) spanning from 2000 to 2017. Extending the investigation until June 2023, we confirm the robustness of these anomalies, with 23 anomalies passing rigorous testing and being chosen for inclusion in our study. Following the study of Qiao (2018), we categorize these anomaly variables into five groups: trade friction, momentum, value and growth, profitability, and intangible assets.

Table 1 presents fundamental information on the 23 stock market anomalies, detailing their abbreviations, full names, and references. Panel A of Table 1 provides basic details on trade friction anomalies, including firm size, idiosyncratic volatility, total volatility, Dimson beta, variation of share turnover, variation of dollar trading volume, turnover-adjusted number of zero daily volume, maximum daily return, and short-term reversal. Panel B focuses on a single momentum anomaly, the prior 24-month momentum. Panel C offers essential insights into value and growth anomalies, such as long-term reversals, price-earnings ratios, quarterly dividend yields, and quarterly sale-to-prices ratios. Panel D presents vital information on profitability anomalies, encompassing change in return on equity, change in return on assets, and quarterly sales growth. Finally, Panel E reports key data on intangible asset anomalies, including the effect of quarterly R&D expenses on market equity and seasonality.

**Table 1**
Basic information on stock market anomalies.

| Abbreviations | Full Names | References |
|---|---|---|
| Panel A. Trading friction | | |
| Size | Firm Size | Banz (1981) |
| idvff1 | Idiosyncratic Volatility per the FF 3-Factor Model | Ang, Hodrick, Xing, and Zhang (2009) |
| idvc1 | Idiosyncratic Volatility per the CAPM | Ali, Hwang, and Trombley (2003) |
| tv1 | Total Volatility | Ang et al. (2009) |
| betaDM1 | The Dimson Beta | Dimson (1979) |
| vturn1 | Variation of Share Turnover | Chordia, Subrahmanyam, and Anshuman (2001) |
| vdtv1 | Variation of Dollar Trading Volume | Chordia et al. (2001) |
| vdtv6 | Variation of Dollar Trading Volume | Chordia et al. (2001) |
| vdtv12 | Variation of Dollar Trading Volume | Chordia et al. (2001) |
| Lm1–1 | Turnover-adjusted Number of Zero Daily Volume | Liu (2006) |
| mdr1 | Maximum Daily Return | Bali, Cakici, and Whitelaw (2011) |
| srev | Short-Term Reversal | Jegadeesh (1990) |
| Panel B. Momentum | | |
| m24–1 | Prior 24-month Momentum | Jegadeesh and Titman (1993) |
| Panel C. Value and growth | | |
| lrev1 | Long-Term Reversal | De Bondt and Thaler (1985) |
| lrev6 | Long-Term Reversal | De Bondt and Thaler (1985) |
| lrev12 | Long-Term Reversal | De Bondt and Thaler (1985) |
| dpq1 | Quarterly Dividend Yield | Litzenberger and Ramaswamy (1979) |
| spq1 | Quarterly Sales-to-Price Ratio | Barbee, Mukherji, and Raines (1996) |
| Panel D. Profitability | | |
| droe1 | Changes in Return on Equity | Hou, Xue, and Zhang (2015) |
| droa1 | Changes in Return on Assets | Balakrishnan, Bartov, and Faurel (2010) |
| sgq1 | Quarterly Sales Growth | Qiao (2018) |
| Panel E. Intangibles assets | | |
| rdmq1 | Quarterly R&D Expense to Market Equity | Chan, Lakonishok, and Sougiannis (2001) |
| Ra25 | Seasonality | Heston and Sadka (2008) |

This table lists abbreviations, full names, and references for stock market anomalies. The methodology for constructing these anomalies is explained in detail in the Appendix.

Table 2 presents the time-series average returns of stock market anomalies and their significance levels as indicated by the Newey-West *t*-statistic (Newey & West, 1987) across the entire sample period and the study sample period. Anomaly returns are calculated by comparing the performance of the highest and lowest deciles based on the anomaly variable. The Newey-West *t*-statistic measures the statistical significance of this return difference, with a threshold of $|t| = 1.96$ determining anomaly presence. The study period selection is critical for ensuring sufficient data availability for the study variables, a point that will be further discussed.

In Panel A of Table 2, a summary of stock market anomalies from January 2000 to June 2023 is provided. The anomalies are established from January 2000, coinciding with China having over 1000 listed companies, ensuring robust sample sizes per decile. Only anomalies that exhibit consistent significance throughout the full sample period are included. Notable findings include the vdtv1 anomaly with the highest average return at 1.41 % and the dpq1 anomaly with the lowest return at 0.42 %. The idvff1 anomaly boasts the highest Newey-West *t*-statistic at 3.71, while the rdmq1 anomaly registers the lowest at 2.10.

Moving to Panel B of Table 2, summary statistics for stock market anomalies between January 2011 and December 2022, the period of our study, are detailed. This timeframe ensures data availability for all study variables. Ten anomalies show decreasing significance in average returns during this period, leading to their exclusion from analysis. Among the 13 anomalies analyzed, nine are labeled as trading friction anomalies, one as a momentum anomaly (m24–1), and two as profitability anomalies (droe1 and droa1). The remaining anomaly, Ra25, falls under the category of intangible assets. Notable figures include the Lm1–1 anomaly with the highest average return at 1.98 % and the Ra25 anomaly with the lowest at 0.71 %. The Lm1–1 anomaly holds the highest Newey-West *t*-statistic at 3.30, while the m24–1 anomaly has the lowest at 1.99.

### *2.2. Investor attention*

This section focuses on creating the investor attention index. Since investor attention cannot be directly observed, we measure it using proxies. The study utilizes 11 individual proxies to capture it following Chen et al. (2022). We further combine the individual proxies to form a composite index of investor attention in the Chinese stock market. Specifically, this composite index is formulated by merging the 11 indicators using the partial least squares (PLS) methodology, as detailed in the subsequent subsection.

#### *2.2.1. Individual attention proxies*

This subsection provides a detailed description of the individual proxies used to measure investor attention. To address the challenge of directly observing investor attention, this study employs 11 proxies specifically tailored to the unique characteristics of the Chinese stock market. These proxies include measures such as abnormal trading volume, extreme returns, past returns, analyst coverage, change in advertising expenses, mutual fund inflows and outflows, media coverage, web search volume index, nearness to the China Securities Free Float Index (CSCI) 52-week highs, and nearness to CSCI historical highs.

**Table 2**
Summary statistics.

Panel A: Full sample time: Jan 2000 to Jun 2023

| | Size | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|---|
| *r* | 1.10** | 1.27*** | 1.14*** | 0.82** | 0.76** | 1.17*** | 1.41*** | 1.08*** |
| *t* | 2.37 | 3.71 | 3.27 | 2.22 | 2.48 | 2.94 | 3.55 | 2.85 |
| | vdtv12 | Lm1–1 | mdr1 | srev | m24–1 | lrev1 | lrev6 | lrev12 |
| *r* | 1.04*** | 1.22*** | 0.68** | 1.08*** | 0.87** | 0.88** | 0.87** | 0.83** |
| *t* | 2.89 | 2.85 | 2.13 | 2.76 | 2.50 | 2.47 | 2.50 | 2.54 |
| | dpq1 | spq1 | droe1 | droa1 | sgq1 | rdmq1 | Ra25 | |
| *r* | 0.42** | 0.73** | 1.03*** | 0.72** | 0.70*** | 0.68** | 0.74*** | |
| *t* | 2.16 | 2.29 | 3.33 | 2.50 | 2.86 | 2.10 | 3.18 | |

Panel B: Study sample time: Jan 2011 to Dec 2022

| | Size$^{N}$ | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|---|
| *r* | 1.09 | 1.60*** | 1.57*** | 1.19** | 0.93** | 1.51** | 1.40*** | 1.14** |
| *t* | 1.59 | 2.97 | 2.75 | 2.07 | 2.19 | 2.47 | 2.61 | 2.33 |
| | vdtv12 | Lm1–1 | mdr1$^{N}$ | srev$^{N}$ | m24–1 | lrev1$^{N}$ | lrev6$^{N}$ | lrev12$^{N}$ |
| *r* | 1.10** | 1.98*** | 0.80* | 1.19 | 0.98** | 0.89* | 0.74 | 0.62 |
| *t* | 2.34 | 3.30 | 1.68 | 1.95 | 1.99 | 1.80 | 1.56 | 1.40 |
| | dpq1$^{N}$ | spq1$^{N}$ | droe1 | droa1 | sgq1$^{N}$ | rdmq1$^{N}$ | Ra25 | |
| *r* | 0.30 | 0.15 | 1.18*** | 0.86** | 0.45 | 0.21 | 0.71** | |
| *t* | 1.09 | 0.28 | 2.92 | 2.31 | 1.52 | 0.43 | 2.02 | |

Panel A of this table provides time-series average returns and Newey-West *t*-statistics for 23 anomalies from January 2000 to June 2023. Panel B of this table provides time-series average returns and Newey-West *t* statistics for 23 anomalies from January 2011 to December 2022. In Panel B, the anomalies labeled "N" in the upper right corner are non-significant. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Table 3, essential information on the 11 individual attention proxies, including abbreviations, full names, and references, is presented. These proxies are categorized into firm-level and market-level indicators. Firm-level proxies are based on data from listed companies, while market-level proxies are calculated as the equal-weighted average of individual attention proxies for all firms. Firm-level proxies consist of various measures, such as abnormal trading volume, extreme returns, past returns, analyst coverage, and more. Market-level proxies focus on the proximity of stocks to CSCI 52-week and historical highs.

Table 4 reports the definitions of individual attention proxies. Specifically, the attention proxies including $A^{AVol}$, $A^{ERet}$, $A^{PRet}$, $A^{\#AC}$, $A^{CAD}$, $A^{Inflow}$, $A^{Outflow}$, $A^{52wH}$, and $A^{HisH}$ are constructed using the basic data through the definitions. Additionally, proxies for media coverage ($A^{Media}$) and web search volume index ($A^{WSVI}$) are directly sourced from the China Research Data Service (CRDS) platform. The $A^{Media}$ index originates from the Financial News Database of Chinese Listed Companies, a leading big data repository using AI algorithms to compile and analyze financial news. The $A^{WSVI}$ is sourced from the Web Search Volume Index of Chinese Listed Companies, systematically organizing web search data linked to listed companies in China since 2011.

**Table 3**
Basic information on individual attention proxies.

| Abbreviations | Full Names | References |
|---|---|---|
| Panel A: Firm-Level Individual Attention Proxies | | |
| $A^{AVol}$ | Abnormal trading volume | Barber and Odean (2008) |
| $A^{ERet}$ | Extreme returns | Barber and Odean (2008) |
| $A^{PRet}$ | Past returns | Aboody et al. (2010) |
| $A^{\#AC}$ | Analyst coverage | Hirshleifer and Teoh (2003) |
| $A^{CAD}$ | Change in advertising expenses | Lou (2014) |
| $A^{Inflow}$ | Mutual fund inflows | Chen et al. (2022) |
| $A^{Outflow}$ | Mutual fund outflows | Chen et al. (2022) |
| $A^{Media}$ | Media coverage | Barber and Odean (2008) |
| $A^{WSVI}$ | Web Search Volume Index | Que and Zhang (2021) |
| Panel B: Market-Level Individual Attention Proxies | | |
| $A^{52wH}$ | Nearness to the CSCI 52-week highs | Li and Yu (2012) |
| $A^{HisH}$ | Nearness to the CSCI historical highs | Li and Yu (2012) |

This table reports basic information on 11 individual attention proxies, including abbreviations, full names, and references.

**Table 4**
Definition of individual attention proxies.

| Indicators | Definition |
|---|---|
| $A^{AVol}$ | The ratio of a stock's volume at the end of each month to its average monthly volume over the previous 12 months |
| $A^{ERet}$ | The ratio of the monthly return of the stock at the end of each month to the average monthly return for the previous 12 months |
| $A^{PRet}$ | Stock's cumulative monthly return over the past 12 months |
| $A^{\#AC}$ | Total number of analysts who predict a stock's earnings per share for the next year in one month |
| $A^{CAD}$ | Changes in the logarithmic value of the firm's advertising costs from year *t* - 1 to year t |
| $A^{Inflow}$ | Monthly subscriptions to the Fund |
| $A^{Outflow}$ | Monthly redemptions from the Fund |
| $A^{Media}$ | Monthly media coverage (available from CRDS) |
| $A^{WSVI}$ | Monthly web search volume index (available from CRDS) |
| $A^{52wH}$ | The ratio of the current month's CSCI to its maximum value over the last 52 weeks (12 months) |
| $A^{HisH}$ | The ratio of the current month's CSCI to its historical maximum |

This table provides the construction method for the individual attention proxies. Based on the definitions given in the right column of the table, we can obtain the firm-level attention proxies. We can then compute the equal weights of the attention proxies of all firms to obtain the market-level attention proxies.

In Table 5, the statistics for the 11 individual attention proxies are presented, including minimum, median, maximum, skewness, and kurtosis. Standardization is applied to all attention proxies to achieve a mean of 0 and variance of 1. The sample data covers the period from January 2011 to December 2022, chosen for data alignment purposes. Notably, all attention indicator series have minimum values below zero. Moreover, the skewness of the majority of attention indicators is positive, except for indicators $A^{ERet}$ and $A^{52wH}$, while most indicators exhibit kurtosis values exceeding 3, indicating right-skewed and peaked distributions.

**Table 5**
Summary statistics for 11 individual attention proxies.

| | Min | Median | Max | Skewness | Kurtosis | Sample Period |
|---|---|---|---|---|---|---|
| $A^{AVol}$ | −0.50 | −0.17 | 11.90 | 9.30 | 98.84 | 1991.12–2023.06 |
| $A^{ERet}$ | −14.00 | 0.00 | 6.17 | −6.28 | 117.33 | 1992.01–2023.06 |
| $A^{PRet}$ | −2.32 | −0.14 | 3.97 | 1.02 | 4.64 | 1992.01–2023.06 |
| $A^{\#AC}$ | −2.06 | −0.08 | 2.60 | 0.36 | 2.40 | 2010.01–2022.12 |
| $A^{CAD}$ | −0.42 | −0.27 | 3.30 | 3.00 | 10.01 | 2011.01–2022.12 |
| $A^{Inflow}$ | −1.07 | −0.37 | 3.26 | 1.85 | 6.07 | 2004.07–2023.06 |
| $A^{Outflow}$ | −0.91 | −0.31 | 3.80 | 1.86 | 6.74 | 2004.07–2023.06 |
| $A^{Media}$ | −1.64 | 0.13 | 3.93 | 0.16 | 2.86 | 2005.01–2022.12 |
| $A^{WSVI}$ | −1.13 | −0.23 | 5.45 | 3.15 | 14.91 | 2011.01–2023.06 |
| $A^{52wH}$ | −6.73 | 0.03 | 6.92 | −0.06 | 23.20 | 2006.12–2023.06 |
| $A^{HisH}$ | −2.95 | −0.02 | 4.81 | 0.55 | 6.39 | 2006.12–2023.06 |

This table reports the minimum, median, maximum, skewness, and kurtosis of the 11 individual attention proxies. All variables of attention are standardized with a mean of 0 and a variance of 1. CSCI stands for China Securities Free Float Index.

Table 6 displays the correlations between the 11 different attention proxies. The correlation coefficients range from −0.36 to 0.88. Notably, the correlation between $A^{Inflow}$ and $A^{Outflow}$ is the highest, as expected. Additionally, indicators $A^{Media}$ and $A^{WSVI}$ from CRDS exhibit a relatively high correlation. The diverse correlations among the other indicators imply that the proxies reflect both shared and unique aspects of investor attention. Relying solely on these individual proxies may not offer a comprehensive understanding of investor attention in the Chinese stock market. In the following sections, we will elaborate on the construction method of composite attention indicators.

### 2.2.2. *Composite index of investor attention*

In contrast to previous studies focusing on a single proxy, we adopt a collection of individual attention proxies, following Chen et al. (2022). Recognizing that investor attention is a latent variable represented by proxies, it is beneficial to isolate the central aspect of true investor attention by refining the data from the 11 proxies. The sample period extends from January 2011 to December 2022.

We propose an explanatory model based on investor attention as follows:

$$r_t = \alpha + \beta A_t^* + \varepsilon_t \tag{1}$$

where $r_t$ is the realized stock market anomaly returns at time $t$, $A_t^*$ is the real but unobservable investor attention at time $t$, and $\varepsilon_t$ is a noise term that is unexplained and unrelated to $A_t^*$. The model in Eq. (1) implies that the real investor attention, $A_t^*$, is correlated with the current anomaly returns.

To efficiently estimate $A_t^*$, we utilize the PLS method to analyze investor attention's effect on anomalies. This method extracts $A_t^*$ from individual attention proxies based on their covariance with the anomaly returns and selects the optimal linear combination of individual attention proxies for an explanation. The PLS can be implemented in the following two steps.

As the first step, we run a time series regression for each attention proxy using the realized anomaly returns $r_t$ in month $t$,

$$A_{i,t} = \alpha_0 + \beta_i r_t + \varepsilon_{i,t} \tag{2}$$

where $A_{i,t}(i = 1, \dots, 11)$ is the data of the $i$-th individual attention proxy in month $t$, $r_t$ is the return of the market anomalies in month $t$. The coefficient $\beta_i$ reflects the sensitivity of the attention proxy $A_{i,t}$ to the investor's attention $A_t^*$ measured by the current anomaly returns $r_t$. The given Eq. (1) shows that the anomaly returns $r_t$ is driven by $A_t^*$. It means that the attention proxy $A_{i,t}$ is associated with the explainable portion of the anomaly returns, independent of the unexplained error. Thus, the coefficient $\beta_i$ approximates the dependence of each attention proxy on the investor's true attention $A_t^*$.

The second regression step is a cross-sectional regression for each time $t$,

$$A_{i,t} = c_t + A_t^{PLS}\widehat{\beta}_i + v_{i,t} \tag{3}$$

where $\widehat{\beta}_i$ is the load estimated in the regression of Eq. (2), $A_t^{PLS}$ is the PLS attention measure in month $t$. To estimate how $A_{i,t}$ depends on $A_t^*$, the regression slope from the first step provides an approximate estimate, $\widehat{\beta}_i$. To extract the relevant $A_t^*$ for interpretation, the PLS method is downscaled using the anomaly returns in month $t$ while discarding irrelevant components.

To empirically investigate, we analyze data spanning from January 2011 to December 2022 to calculate the composite index of investor attention and assess its interpretability within the sample. In the first step, we estimate the load ($\beta_i$) of the anomaly returns on each of the 11 attention proxies, namely $A^{AVol}$, $A^{ERet}$, $A^{Pret}$, $A^{52wH,}$ $A^{HisH}$, $A^{\#\ AC}$, $A^{CAD}$, $A^{Inflow}$, $A^{Outflow}$, $A^{Media}$, and $A^{WSVI}$. Subsequently, we estimate $A_t^{PLS}$ based on the loadings $\beta_i$. Thus, we can obtain the monthly PLS-based investor attention composite index $A_t^{PLS}$ linking to the specific anomaly returns.

**Table 6**
Correlation among 11 individual attention proxies.

| | $A^{ERet}$ | $A^{PRet}$ | $A^{\#AC}$ | $A^{CAD}$ | $A^{Inflow}$ | $A^{Outflow}$ | $A^{Media}$ | $A^{WSVI}$ | $A^{52wH}$ | $A^{HisH}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $A^{AVol}$ | −0.02 | −0.05 | −0.06 | −0.05 | −0.05 | −0.06 | 0.02 | 0.00 | −0.08 | −0.03 |
| $A^{ERet}$ | | 0.02 | −0.05 | 0.00 | −0.10 | −0.06 | −0.11 | −0.08 | 0.05 | −0.01 |
| $A^{PRet}$ | | | −0.32 | −0.36 | 0.62 | 0.61 | 0.39 | 0.74 | 0.02 | −0.02 |
| $A^{\#AC}$ | | | | 0.39 | −0.21 | −0.13 | −0.06 | −0.21 | −0.15 | −0.09 |
| $A^{CAD}$ | | | | | −0.23 | −0.22 | −0.04 | −0.15 | 0.02 | 0.01 |
| $A^{Inflow}$ | | | | | | 0.88 | 0.42 | 0.68 | 0.04 | 0.10 |
| $A^{Outflow}$ | | | | | | | 0.42 | 0.67 | 0.08 | 0.12 |
| $A^{Media}$ | | | | | | | | 0.41 | 0.03 | 0.02 |
| $A^{WSVI}$ | | | | | | | | | 0.04 | 0.05 |
| $A^{52wH}$ | | | | | | | | | | 0.73 |

This table displays the two-by-two correlations of the 11 individual attention proxies. All variables of attention are standardized with a mean of 0 and a variance of 1. CSCI stands for China Securities Free Float Index.

## 3. Empirical results

### 3.1. Results based on investor attention constructed by the PLS method

In this section, we begin by investigating the relationship between investor attention and market anomalies. The benchmark model is defined as follows:

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t} IA_{i,t}^{PLS} + \varepsilon_{i,t} \tag{4}$$

where $r_{i,t}$ is the return of stock market anomaly $i$ in month $t$. Anomaly $i$ denotes idvff1, idvc1, tv1, betaDM1, vturn1, vdv1, vdtv6, vdtv12, Lm1–1, m24–1, droe1, droa1, and Ra25, respectively. The composite index of investor attention denoted as $IA_{i,t}^{PLS}$, is developed using the PLS method as explained in subsection 2.2.2. The coefficient $\beta_{i,t}$ signifies the influence of investor attention on simultaneous anomaly returns.

In Table 7, the regression findings between investor attention and stock market anomalies are displayed utilizing Eq. (4). These results reveal a significant positive relationship between investor attention and concurrent market anomalies, as indicated by coefficients exceeding 0. Notably, idvc1 demonstrates the highest Newey-West $t$-statistic of 9.28, while Ra25 shows the lowest at 1.97. These results suggest that anomaly returns provide insights into investor attention during the same period, highlighting a positive impact of investor attention on anomaly returns. Building on the research by Jiang, Liu, Peng, and Wang (2021), we propose that increased investor attention leads to heighten investor overreaction, magnifying mispricing. Yet, limitations in arbitrage impede prompt corrections, consequently enhancing the profitability of market anomalies.

### 3.2. Controlling for the Fama-French three factors

Our prior analysis demonstrates a positive relationship between investor attention and anomaly returns in China's stock market. We continue to argue that this effect holds after controlling for some well-known control variables. To do this, we will first select market, size, and value factors from the Fama-French three-factor model (Fama & French, 1993). Then, the model that examines the impact of investors' attention on market anomalies will be extended by:

$$r_{i,t} = \alpha + \beta_{i,t} IA_{i,t}^{PLS} + \beta_{MKT} MKT_t + \beta_{SMB} SMB_t + \beta_{HML} HML_t + \varepsilon_{i,t} \tag{5}$$

where MKT, SMB, and HML are market, size, and value factors constructed based on mock portfolios. The data are obtained from the Resset database. The coefficient $\beta_{i,t}$ measures the influence of investor attention on contemporaneous anomaly returns after controlling for market, size, and value factors.

Table 8 presents the regression results for examining the relationship between investor attention and anomaly returns with the consideration of Fama-French's three factors. We can observe a significant positive correlation between anomaly returns and investor attention, with exceptions for betaDM1 and Lm1–1 anomalies. These findings are consistent with those from Table 7, which did not involve market, size, and value control variables. The results highlight anomaly returns influenced by investor attention, independent of MKT, SMB, and HML factors.

### 3.3. Controlling for investor sentiment

Investor sentiment is another critical factor in investor behavior that has drawn significant attention from scholars investigating its relationship with stock market returns. A notable impact of investor sentiment on anomaly returns has been found (Stambaugh et al., 2012). Inspired by this, we aim to incorporate investor sentiment control in examining the influence of investor attention on market anomalies. The specific model setup is as follows:

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t}^{IA} IA_{i,t}^{PLS} + \beta_{i,t}^{IS} IS_{i,t}^{PLS} + \varepsilon_{i,t} \tag{6}$$

where $IA_{i,t}^{PLS}$ represents the investor attention index and $IS_{i,t}^{PLS}$ denotes the investor sentiment. Following the methodology of Yi and Mao (2009), we use six indicators to construct a composite index of investor sentiment by the PLS method. These indicators include the fund discount rate, last month's trading volume, number of IPOs, first-day IPO returns, number of new investor accounts opened last month, and consumer confidence index. The data for these six indicators are obtained from CSMAR.

Table 9 displays the regression outcomes of investor attention on anomaly returns following the inclusion of investor sentiment. Compared to Table 7, all anomalies, except for the Ra25, display a significant and positive correlation with investor attention. Furthermore, it is noted that investor sentiment exerts a noteworthy positive influence solely on anomalies vturn1, vdtv1, vdtv6, vdtv12, and m24–1. The finding indicates that the effect of investor attention on market anomalies remains robust even when considering investor sentiment variables. We can also deduce that investor attention and sentiment can be viewed as isolated behavioral biases influencing stock markets. Additionally, Table 9 illustrates that investor attention's impact on anomalies outweighs investor sentiment.

### 3.4. Results based on investor attention and anomaly returns constructed by bivariate grouping

In our previous analysis, we provide evidence of the significant positive correlation between investor attention and stock market anomaly returns. Specifically, higher investor attention corresponds to increased anomaly returns, as indicated by our time series regression methodology. To confirm our findings, we further conduct a cross-sectional analysis incorporating the web search volume index as the attention proxy for bivariate grouping with the anomaly variable. The $A^{WSVI}$ index contains detailed information linking stock codes, company abbreviations, and full company names of Chinese listed firms, serving as a reliable measure of individual attention at the firm level and showing advantages in performing the cross-sectional analysis.

To be specific, we begin by dividing the sample stocks into 10 groups based on anomaly variables, then categorize them into low, middle, and

**Table 7**
Regression results for investor attention constructed by the PLS method and anomaly returns.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| $\beta_{i,t}$ | 0.31*** | 0.30*** | 0.26*** | 0.14*** | 0.18*** | 0.07*** | 0.07*** |
| $t$ | 9.14 | 9.28 | 7.55 | 3.97 | 6.56 | 2.80 | 2.89 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}$ | 0.07*** | 0.23*** | 0.09*** | 0.07*** | 0.08*** | 0.03** | |
| $t$ | 2.98 | 7.09 | 3.72 | 5.60 | 5.74 | 1.97 | |

This table reports the regression results for investor attention constructed by the PLS method and anomaly returns. The "$\beta_{i,t}$" in this table indicates the extent to which investor attention affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 8**
Controlling for the Fama-French three factors.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| $\beta_{i,t}$ | 0.18*** | 0.18*** | 0.13** | −0.02 | 0.09** | 0.02* | 0.01* |
| $t$ | 4.03 | 4.05 | 2.07 | −0.33 | 2.26 | 1.74 | 1.64 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}$ | 0.01* | 0.08 | 0.05*** | 0.05*** | 0.07*** | 0.04** | |
| $t$ | 1.64 | 1.52 | 2.61 | 3.24 | 4.29 | 2.26 | |

This table reports the results after controlling for the Fama-French three factors. The "$\beta_{i,t}$" in this table indicates the extent to which investor attention affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 9**
Controlling for investor sentiment.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| $\beta_{i,t}^{IA}$ | 0.3*** | 0.29*** | 0.25*** | 0.14*** | 0.17*** | 0.06** | 0.03* |
| $t$ | 7.18 | 7.05 | 7.10 | 3.76 | 4.99 | 2.32 | 1.85 |
| $\beta_{i,t}^{IS}$ | 0.01 | 0.01 | 0.02 | 0.01 | 0.05** | 0.03* | 0.02** |
| $t$ | 0.39 | 0.27 | 0.71 | 1.31 | 2.08 | 1.95 | 1.98 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}^{IA}$ | 0.06*** | 0.22*** | 0.07*** | 0.07*** | 0.07*** | 0.03 | |
| $t$ | 2.63 | 5.86 | 3.19 | 4.39 | 4.54 | 1.57 | |
| $\beta_{i,t}^{IS}$ | 0.04*** | 0.03 | 0.04*** | 0.00 | 0.01 | 0.01 | |
| $t$ | 2.69 | 1.36 | 2.86 | 0.42 | 0.60 | 0.57 | |

This table reports the results after controlling for investor sentiment. The "$\beta_{i,t}^{IA}$" in this table indicates the extent to which investor attention affects market anomaly returns. The "$\beta_{i,t}^{IS}$" in this table indicates the extent to which investor sentiment affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

high web search volume index groups using quartiles at 30 % and 70 %. This created 30 subgroups, each comprising low, middle, and high web search volume index subgroups with 10 anomaly variable portfolios ranging from low to high. Anomaly returns can be calculated for each of the three web search volume index groupings. Using bivariate grouping instead of univariate grouping allows for a more comprehensive analysis of the relationship between investor attention and anomalies in the Chinese stock market. This method enables the segregation of investor attention levels to delve deeper into understanding how they influence anomalies in the cross-section.

Table 10 illustrates the bivariate grouping results of the attention proxy and anomalies. Anomalies idvc1, tv1, vturn1, Lm1–1, droe1, and droa1 exhibit rising returns corresponding to an increase in the web search volume index across all three groups. When comparing the high and low web search volume index groups, anomalies, except betaDM1, demonstrate significantly higher returns in the high web search volume index group. The most substantial discrepancy in anomaly returns between the high and low web search volume indices is observed in vturn1 (H-L = 2.00 %), while the smallest differences are evident in droe1 and Ra25 (H-L = 0.27 %). Through the bivariate grouping test, a positive correlation is identified between investor attention and anomaly returns.

**Table 10**
Anomaly returns based on bivariate grouping of web search volume indices and anomalies.

| $A^{WSVI}$→ | Low | Middle | High | H-L |
|---|---|---|---|---|
| idvff1 | 1.30 | 1.62 | 1.59 | 0.29 |
| idvc1 | 1.14 | 1.29 | 1.68 | 0.54 |
| tv1 | 0.11 | 0.72 | 1.30 | 1.19 |
| betaDM1 | 1.29 | 0.66 | 0.94 | −0.35 |
| vturn1 | 0.22 | 0.86 | 2.22 | 2.00 |
| vdtv1 | 0.55 | 1.47 | 0.99 | 0.44 |
| vdtv6 | 0.34 | 1.17 | 0.78 | 0.44 |
| vdtv12 | 0.40 | 1.28 | 0.77 | 0.37 |
| Lm1–1 | 0.86 | 1.68 | 2.24 | 1.38 |
| m24–1 | 0.67 | 0.65 | 0.99 | 0.32 |
| droe1 | 0.87 | 0.99 | 1.19 | 0.32 |
| droa1 | 0.62 | 0.75 | 0.89 | 0.27 |
| Ra25 | 0.24 | 0.55 | 0.51 | 0.27 |

Bivariate independent grouping is performed for all sample stocks. First, the sample stocks are divided into 10 groups based on anomaly variables. Then, the sample stocks are divided into low, middle, and high groups based on the web search volume index, using 30 % and 70 % as the quartiles. Finally, 30 subgroups are obtained. Each low, middle, and high web search volume index subgroup has 10 subgroups of anomaly variables, ranging from low to high. The anomaly returns are constructed for the three web search volume index groupings.

Building upon the bivariate grouping findings and prior research, we explore whether the investor attention composite index, derived from 11 attention proxies, influences anomaly returns differently across various web search volume index groups. We posit that the investor attention composite index has a more pronounced explanatory effect on anomaly returns in the high web search volume index group. To assess this hypothesis, we substitute the left-hand side variables of Eq. (4) with anomaly returns categorized by bivariate grouping, encompassing low, medium, and high Web search volume index groups.

The results presented in Table 11 focus on investor attention and anomaly returns organized by bivariate grouping. The coefficient values of investor attention are indicated on the left side of Table 11, while the corresponding Newey-West $t$ statistics are provided on the right side. Analysis of the coefficient values reveals that, except for the betaDM1 anomaly, investor attention coefficients for other anomalies consistently rise with the web search volume index. Notably, significant Newey-West $t$ statistics are observed for eight anomalies (idvff1, idvc1, tv1, vturn1, vdtv1, Lm1–1, droe1, and droa1). Collectively, these results suggest that anomaly returns linked to a high web search volume index offer more insights into investor attention, highlighting a stronger explanatory power of investor attention. The robustness of the findings in Table 11 reinforces our hypothesis validity, affirming the superior explanatory role of investor attention on anomaly returns linked to a high web search

**Table 11**
Results based on investor attention and anomaly returns constructed by bivariate grouping.

| $A^{WSVI}$→ | Low | Middle | High | Low | Middle | High |
|---|---|---|---|---|---|---|
| | $\beta$ | | | $t(\beta)$ | | |
| idvff1 | 0.04*** | 0.05*** | 0.10*** | 2.89 | 4.51 | 5.46 |
| idvc1 | 0.04*** | 0.05*** | 0.09*** | 2.68 | 4.76 | 4.69 |
| tv1 | 0.02 | 0.05*** | 0.06*** | 1.26 | 3.51 | 3.47 |
| betaDM1 | 0.01 | −0.00 | −0.01 | 0.50 | −0.28 | −0.47 |
| vturn1 | 0.01 | 0.03*** | 0.05*** | 0.74 | 2.69 | 3.13 |
| vdtv1 | 0.01 | 0.01 | 0.04*** | 0.32 | 0.42 | 2.75 |
| vdtv6 | −0.02 | −0.00 | 0.02 | −1.62 | −0.19 | 1.44 |
| vdtv12 | −0.02* | −0.00 | 0.02 | −1.79 | −0.15 | 1.30 |
| Lm1–1 | 0.03* | 0.04** | 0.07*** | 1.75 | 2.11 | 2.64 |
| m24–1 | 0.01 | 0.02 | 0.02 | 1.05 | 1.42 | 1.30 |
| droe1 | 0.03* | 0.03* | 0.05** | 1.80 | 1.91 | 2.09 |
| droa1 | 0.02 | 0.04** | 0.05* | 1.39 | 2.23 | 1.82 |
| Ra25 | 0.00 | −0.00 | 0.01 | 0.24 | −0.08 | 0.34 |

This table reports the results based on investor attention and anomaly returns constructed by bivariate grouping. Abbreviations for the 13 anomalies are shown on the far left side of this table. On the left-hand side, the "Low", "Middle", and "High" represent the results of regressions of investor attention on the anomalies returns for the low, middle, and high attention groups, respectively. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

volume index.

## 4. Extension analyses

### 4.1. Results based on lagged investor attention

A notable positive correlation between investor attention and current anomaly returns is observed in the preceding section, prompting us to explore the potential link to future anomaly returns. To address this, we conduct a test to examine how investor attention impacts future anomaly returns. Mathematically, the model used for this analysis is given as follows:

$$r_{i,t+1} = \alpha_{i,t} + \beta_{i,t} IA_{i,t}^{PLS} + \varepsilon_{i,t+1} \tag{7}$$

where $r_{i,t+1}$ represents the return of stock market anomaly $i$ in month $t+1$, and $IA_{i,t}^{PLS}$ denotes the investor attention composite index constructed by the PLS method in month $t$.

Table 12 presents the results of a regression analysis that examines investor attention's impact on future anomaly returns in the Chinese stock market. We focus on the immediate impact (next 1 period) due to the decreasing importance of investor attention on anomalies over longer periods. The findings in Table 12 highlight a notable influence of investor attention on market anomalies in the following month. Notably, the coefficient values vary, with the Ra25 anomaly having the highest coefficient ($\beta = 0.07$) and the m24–1 anomaly showing the lowest coefficient ($\beta = 0.02$). Importantly, all coefficients are statistically significant at the 10 % level.

In alignment with Section 3.2 and Section 3.3, this subsection delves into conducting additional robustness tests on the established findings. Panel A of Table 13 presents the results of how investor attention impacts stock market anomaly returns when factoring in the three Fama-French factors. A comparison with Table 12, which lacks control for these factors, reveals that while the significance of investor attention on the vdtv1 anomaly diminishes, largely due to the influence of the Fama-French factors, the effects of investor attention on other anomalies remain significant, albeit with reduced strength in the coefficients. Moreover, Panel B illustrates the impact of investor attention on stock market anomaly returns after incorporating investor sentiment. Interestingly, unlike the findings in Table 12, the significance of investor attention on the m24–1 anomaly diminishes. Nevertheless, the notable effects on other anomalies persist. This observation indicates that investor sentiment does not undermine the impact of investor attention on anomaly returns.

### 4.2. An alternative method for constructing a composite index of investor attention

In Section 3, we employ the investor attention composite index created through the PLS technique. Now, we aim to construct a composite investor attention index based on an alternative method and further examine its relationship with anomaly returns in the Chinese stock market. Specifically, we employ another widely used dimensionality reduction method, namely the scaled principal component analysis (sPCA) proposed by Huang, Jiang, Li, Tong, and Zhou (2022). PCA focuses on capturing the most variance in explanatory factors, disregarding the target variable, making it an unsupervised learning approach for dimensionality reduction. In contrast, sPCA integrates target information to guide the dimensionality reduction process.

The implementation of the sPCA technique involves two steps. First, a collection of scaled attention explanatory factors is generated ($\beta_1 A_{i,t}$, …, $\beta_N A_{N,t}$), where each scaling factor $\beta_i$ ($i = 1, …, N$) represents the slope obtained from regressing the anomaly returns of the current period on the $i$-th attention proxy ($IA_{i,t}$).

$$r_{i,t} = \alpha_{i,t} + \beta_{i,t} IA_{i,t} + \varepsilon_{i,t} \tag{8}$$

In the second step of the process, the PCA method is applied to the set of scaled explanatory factors ($\beta_1 A_{i,t}$, …, $\beta_N A_{N,t}$). The results in the creation of the principal component specifically focus on investor attention, referred to as $IA^{sPCA}$, which is established through the sPCA method. In summary, sPCA conducts PCA on the scaled attention proxies instead of the original ones.

Table 14 presents the regression results regarding investor attention constructed by the sPCA method and its impact on anomaly returns. Our analysis indicates that the investor attention index constructed using the sPCA method significantly influences anomaly returns in the stock market. Specifically, all regression coefficients are significantly positive, consistent with the findings from the PLS method. Most regression coefficients for investor attention on market anomaly returns are significant at the 1 % level. In conclusion, the empirical results reaffirm the substantial positive impact of the composite investor attention index on

**Table 12**
Results based on lagged investor attention.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| $\beta_{i,t}$ | 0.04*** | 0.04*** | 0.04*** | 0.04*** | 0.04*** | 0.04** | 0.04** |
| $t$ | 2.84 | 2.66 | 3.60 | 3.26 | 3.95 | 2.00 | 2.13 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}$ | 0.04** | 0.04*** | 0.02* | 0.06*** | 0.05*** | 0.07*** | |
| $t$ | 2.36 | 3.57 | 1.84 | 3.48 | 2.77 | 3.31 | |

This table reports the results based on lagged investor attention. The "$\beta_{i,t}$" in this table indicates the extent to which lagged investor attention affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 13**
Robustness tests based on lagged investor attention.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| Panel A: Control for the Fama-French three factors | | | | | | | |
| $\beta_{i,t}$ | 0.02*** | 0.03*** | 0.03*** | 0.04*** | 0.03*** | 0.01 | 0.01** |
| $t$ | 2.97 | 2.96 | 2.85 | 3.60 | 3.56 | 1.63 | 1.97 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}$ | 0.02** | 0.02*** | 0.01* | 0.05*** | 0.04** | 0.07*** | |
| $t$ | 2.40 | 3.06 | 1.65 | 2.74 | 2.46 | 3.70 | |
| Panel B: Control for investor sentiment | | | | | | | |
| $\beta_{i,t}^{IA}$ | 0.03* | 0.03* | 0.03*** | 0.03*** | 0.04*** | 0.03* | 0.03* |
| $t$ | 1.82 | 1.72 | 3.16 | 2.66 | 3.78 | 1.68 | 1.85 |
| $\beta_{i,t}^{IS}$ | 0.02 | 0.02 | 0.02* | 0.01 | 0.03** | 0.02* | 0.02** |
| $t$ | 1.42 | 1.28 | 1.74 | 0.72 | 2.34 | 1.89 | 1.98 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}^{IA}$ | 0.03** | 0.04*** | 0.01 | 0.04* | 0.03* | 0.07*** | |
| $t$ | 2.08 | 3.72 | 0.49 | 1.92 | 1.79 | 3.45 | |
| $\beta_{i,t}^{IS}$ | 0.02** | 0.03** | 0.06*** | 0.04** | 0.04** | 0.01 | |
| $t$ | 2.00 | 2.27 | 2.84 | 2.24 | 2.09 | 1.12 | |

This table reports the robustness tests based on lagged investor attention. The coefficient "$\beta_{i,t}^{IA}$" indicates the extent to which lagged investor attention affects market anomaly returns, and the coefficient "$\beta_{i,t}^{IS}$" indicates the extent to which lagged investor sentiment affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

**Table 14**
Regression results for investor attention constructed by the sPCA method and anomaly returns.

| | idvff1 | idvc1 | tv1 | betaDM1 | vturn1 | vdtv1 | vdtv6 |
|---|---|---|---|---|---|---|---|
| $\beta_{i,t}$ | 0.82*** | 0.82*** | 0.83*** | 0.82*** | 0.83*** | 0.66** | 0.66** |
| $t$ | 8.00 | 8.04 | 7.43 | 5.33 | 6.48 | 2.45 | 2.41 |
| | vdtv12 | Lm1–1 | m24–1 | droe1 | droa1 | Ra25 | |
| $\beta_{i,t}$ | 0.66** | 0.83*** | 0.84*** | 0.64*** | 0.62*** | 0.71* | |
| $t$ | 2.40 | 8.14 | 2.68 | 3.19 | 2.64 | 1.94 | |

This table reports the regression results for investor attention constructed by the sPCA method and anomaly returns. The "$\beta_{i,t}$" in this table indicates the extent to which investor attention constructed by the sPCA method affects market anomaly returns. The "$t$" in this table indicates Newey-West $t$ statistics. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

anomaly returns in the Chinese stock market.

## 5. Conclusion

This study investigates the influence of investor attention on anomalies in the Chinese stock market. We employ 11 individual attention proxies to create an investor attention composite index using the PLS method for dimensionality reduction. Our analysis uncovers a significant positive connection between investor attention and market anomalies in the Chinese stock market. A surge in investor attention triggers an immediate increase in anomaly strategy returns. To validate our findings, we conduct robustness tests. Incorporating the Fama-French three factors reveals a strong positive link between market anomaly returns and investor attention. Additionally, introducing the investor sentiment variable, akin to investor attention, reaffirms the substantial impact of investor attention on market anomalies compared to investor sentiment.

In our extension analyses, we address two key aspects. First, we explore the relationship between investor attention and future anomaly returns. Our findings indicate that investor attention impacts anomaly returns in the subsequent month, with minimal influence over longer periods. Second, we examine if creating an investor attention composite index through alternative methods yields consistent findings. We confirm that the investor attention composite index derived from scaled PCA demonstrates a notable positive effect on market anomalies, aligning with our main findings.

## Declaration of competing interest

None.

## Data availability

Data will be made available on request.

## Acknowledgments

This work is supported by the financial support from the National Natural Science Foundation of China [71901122, 72303098], the Humanities and Social Science Fund of Ministry of Education of China [24YJCZH334], the General Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province [2023SJYB0023], and the Fundamental Research Funds for the Central Universities [30922010105].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.irfa.2024.103775.

## 6 references


Aboody, D., Lehavy, R., & Trueman, B. (2010). Limited attention and the earnings announcement returns of past stock market winners. *Review of Accounting Studies, 15*(2), 317–344.

Abreu, D., & Brunnermeier, M. K. (2002). Synchronization risk and delayed arbitrage. *Journal of Financial Economics, 66*(2–3), 341–360.

Ali, A., Hwang, L. S., & Trombley, M. A. (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics, 69*(2), 355–373.

Andrei, D., Friedman, H., & Ozel, N. B. (2023). Economic uncertainty and investor attention. *Journal of Financial Economics, 149*(2), 179–217.

Ang, A., Hodrick, R. J., Xing, Y., & Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further U.S. evidence. *Journal of Financial Economics, 91*(1), 1–23.

Atilgan, Y., Bali, T. G., Demirtas, K. O., & Gunaydin, A. D. (2020). Left-tail momentum: Underreaction to bad news, costly arbitrage and equity returns. *Journal of Financial Economics, 135*(3), 725–753.

Balakrishnan, K., Bartov, E., & Faurel, L. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics, 50*(1), 20–41.

Bali, T. G., Cakici, N., & Whitelaw, R. F. (2011). Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics, 99*(2), 427–446.

Ballinari, D., Audrino, F., & Sigrist, F. (2022). When does attention matter? The effect of investor attention on stock market volatility around news releases. *International Review of Financial Analysis, 82*, Article 102185.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics, 9*(1), 3–18.

Barbee, W. C., Mukherji, S., & Raines, G. A. (1996). Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal, 52*, 56–60.

Barber, B., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies, 21*, 785–818.

Barber, B. M., & Odean, T. (2013). The behavior of individual investors. *Handbook of the Economics of Finance, 2*, 1533–1570.

Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *Review of Financial Studies, 30*(9), 3009–3047.

Chan, L. K. C., Lakonishok, J., & Sougiannis, T. (2001). The stock market valuation of research and development expenditures. *The Journal of Finance, 56*(6), 2431–2456.

Chen, J., Tang, G., Yao, J., & Zhou, G. (2022). Investor attention and stock returns. *Journal of Financial and Quantitative Analysis, 57*(2), 455–484.

Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics, 59*(1), 3–32.

Da, Z., Hua, J., Hung, C.-C., & Peng, L. (2024). Market returns and a tale of two types of attention. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3551662

Da, Z. H. I., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance, 66*(5), 1461–1499.

De Bondt, W. F. M., & Thaler, R. (1985). Does the stock market overreact? *Journal of Finance, 40*(3), 793–805.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *Journal of Finance, 64*(2), 709–749.

Dimson, E. (1979). Risk measurement when shares are subject to infrequent trading. *Journal of Financial Economics, 7*(2), 197–226.

Dong, D., Wu, K., Fang, J., Gozgor, G., & Yan, C. (2022). Investor attention factors and stock returns: Evidence from China. *Journal of International Financial Markets Institutions and Money, 77*, Article 101499.

Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The comovement of investor attention. *Management Science, 63*(9), 2847–2867.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2015). The determinants and consequences of information acquisition via EDGAR. *Contemporary Accounting Research, 32*(3), 1128–1161.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, 33*(1), 3–56.

Han, B., Hirshleifer, D., & Walden, J. (2022). Social transmission bias and investor behavior. *Journal of Financial and Quantitative Analysis, 57*(1), 390–412.

Han, C., & Shi, Y. (2022). Chinese stock anomalies and investor sentiment. *Pacific-Basin Finance Journal, 73*(C). S0927538X22000348.

He, Z., & Manela, A. (2016). Information acquisition in rumor-based bank runs. *Journal of Finance, 71*(3), 1113–1158.

Heston, S. L., & Sadka, R. (2008). Seasonality in the cross-section of stock returns. *Journal of Financial Economics, 87*(2), 418–445.

Hirshleifer, D., Hsu, P.-H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of Financial Economics, 107*(3), 632–654.

Hirshleifer, D., Lim, S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance, 64*(5), 2289–2325.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics, 36*(1), 337–386.

Hou, K., Xue, C., & Zhang, L. (2015). Editor's choice digesting anomalies: An investment approach. *Review of Financial Studies, 28*(3), 650–705.

Hu, Y., Li, X., Goodell, J. W., & Shen, D. (2021). Investor attention shocks and stock co-movement: Substitution or reinforcement? *International Review of Financial Analysis, 73*, Article 101617.

Huang, D., Jiang, F., Li, K., Tong, G., & Zhou, G. (2022). Scaled PCA: A new approach to dimension reduction. *Management Science, 68*(3), 1678–1695.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance, 45*(3), 881–898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance, 48*(1), 65–91.

Jiang, L., Liu, J., Peng, L., & Wang, B. (2021). Investor attention and asset pricing anomalies. *Review of Finance, 26*(3), 563–593.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Karlsson, N., Loewenstein, G., & Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty, 38*(2), 95–115.

Lan, Q., Xie, Y., Mi, X., & Zhang, C. (2024). Post earnings announcement drift: A simple earnings surprise measure, the medium effect of investor attention and investing strategy. *International Review of Financial Analysis, 95*, Article 103460.

Lee, C. M. C., Ma, P., & Wang, C. C. Y. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics, 116*(2), 410–431.

Li, J., & Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability. *Journal of Financial Economics, 104*(2), 401–419.

Lin, P. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis, 40*, 307–329.

Litzenberger, R. H., & Ramaswamy, K. (1979). The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence. *Journal of Financial Economics, 7*(2), 163–195.

Liu, W. (2006). A liquidity-augmented capital asset pricing model. *Journal of Financial Economics, 82*(3), 631–671.

Lou, D. (2014). Attracting investor attention through advertising. *Review of Financial Studies, 27*(6), 1797–1829.

Meng, Y., Li, X., & Xiong, X. (2024). Information shocks and short-term market overreaction: The role of investor attention. *International Review of Financial Analysis, 93*, Article 103219.

Meshcheryakov, A., & Winters, D. B. (2022). Retail investor attention and the limit order book: Intraday analysis of attention-based trading. *International Review of Financial Analysis, 81*, Article 101627.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica, 55*, 703–708.

Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics, 80*(3), 563–602.

Qiao, F. (2018). Replicating anomalies in China. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3263990

Que, J., & Zhang, X. (2021). Money chasing hot industries? Investor attention and valuation of venture capital backed firms. *Journal of Corporate Finance, 68*, Article 101949.

Que, J., Qin, X., & Zhang, Q. (2022). Institutional investor attention and venture capital valuation: evidence from China. *Applied Economics Letters, 30*, 1–7.

Shi, Y., Wang, H., Xia, Y., & Zhen, H. (2023). Mispricing and anomalies in China. *Pacific-Basin Finance Journal, 79*, Article 102038.

Shue, K., & Townsend, R. R. (2021). Can the market multiply and divide? Non-proportional thinking in financial markets. *Journal of Finance, 76*(5), 2307–2357.

Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics, 104*(2), 288–302.

Szczygielski, J. J., Charteris, A., Bwanya, P. R., & Brzeszczyński, J. (2024). Google search trends and stock markets: Sentiment, attention or uncertainty? *International Review of Financial Analysis, 91*, Article 102549.

Yi, Z., & Mao, N. (2009). Measuring investor sentiment in the Chinese stock market: The construction of the CICSI. *Journal of Financial Research, 11*, 174–184 (in Chinese).

Zhang, Y., Chen, Y., Wu, Y., & Zhu, P. (2022). Investor attention and carbon return: evidence from the EU-ETS. *Economic Research-Ekonomska Istraživanja, 35*(1), 709–727.

Zhang, X., Li, G., Li, Y., Zou, G., & Wu, J. G. (2023). Which is more important in stock market forecasting: Attention or sentiment? *International Review of Financial Analysis, 89*, Article 102732.

Zhu, P., Zhang, X., Wu, Y., Zheng, H., & Zhang, Y. (2021). Investor attention and cryptocurrency: Evidence from the Bitcoin market. *PLOS ONE, 16*(2), Article e0246331.