A Tale of Two News-implied Linkages: Information Structure, Processing Costs and Cross-firm Predictability

信息结构，处理成本与股票间收益率可预测性

Shuyi Ge [*], Shaoran Li [†], Hanyu Zheng [‡]

April 11, 2024

Abstract

This paper decomposes news-implied linkages into two types: leader-follower links (LF) and peer links (PE), based on people's reading and information-processing habits. We explore how the structure of information impacts processing costs and subsequently leads to market outcomes by examining momentum spillover effects via these distinct linkage types. Our findings indicate that the information structure of leader-follower links is more readily comprehensible to investors than peer linkages. We provide empirical evidence of this by demonstrating faster attention spillover from leader to follower than among peer firms, using Baidu search data. Furthermore, we document that due to the lower information processing cost, information transmits through the leader-follower linkages more quickly, leading to a weaker momentum spillover effect compared to the more complex and less easily perceivable peer links.

Keywords: Information processing costs; Limited attention; Cross-firm momentum; News; Big textual data; Linked firms; Attention spillover; Return predictability.

JEL Classification: G11, G12, G14

[*] Assistant Professor, The School of Finance, Nankai University. Email: sg751_shuyige@outlook.com

[†] Assistant Professor, The School of Economics, Peking University. Email: lishaoran@pku.edu.cn

[‡] PhD candidate, Faculty of Business and Economics, The University of Hong Kong. Email: nickzhy@outlook.com

1 Introduction

The literature widely acknowledges journalists' role in repackaging, republishing, and disseminating news to the public (e.g., Huberman and Regev (2001); Bushee et al. (2010); Engelberg and Parsons (2011); Lawrence et al. (2018); Blankespoor et al. (2018)). According to Fama (1970), such public information should be immediately reflected in asset prices. However, investors may rationally disregard information from the news if the cost of processing it outweighs the potential benefits. Utilizing a novel financial news database, we differentiate various cross-firm linkages identified from the news based on different information structures. We then examine how information structure affects processing cost, and how this, in turn, influences market outcomes.

We categorize news-implied linkages into two distinct types: the leader-follower (LF) link and the peer link (PE). The classification is based on the positions of firms within a news item, which naturally results in different levels of information processing costs. During a pre-specified identification window, a leader-follower (LF) link is defined when one stock (the leader) appears in a news item's title, and another stock (the follower) appears in the body of the same news item. A leader-follower link is directed where information flows from more attention-grabbing leaders to followers. In contrast to this directed relationship, we also establish undirected peer connections in news where both entities are of equal importance, which we define as peer (PE) links. The PE link captures the relationship between stocks that have only been mentioned in the body of the same news article. To construct PE links, we first define the news co-mention (CM) link between two stocks if they are co-mentioned in the same sentence of a news item. Then PE links are constructed by excluding all LF links from news co-mention links (CM). This is based on the rationale that if a pair has both CM and LF links during the identification window, they do not exhibit equal importance, which is a prerequisite for PE links. Throughout this paper, we mainly focus on comparing the LF and PE links, which exhibit distinct information structures and entail different information processing costs. Specifically, we hypothesize that the LF link has a clearer information structure and requires lower information processing costs than the PE link.

The definitions and hypothesis align seamlessly with news editing practices and readers' reading habits. In news editing, stocks with compelling selling points or undergoing noteworthy events are more likely to be featured in the article's title (Hu and Härdle, 2021). Furthermore, readers tend to focus on headlines and quickly skim through the main content.[1] According to a Microsoft survey conducted in Canada, the average attention span of readers decreased to 8 seconds in 2013, down from 12 seconds in 2000.[2] With less reading time, it's easier for readers to link firms in the title with those in the article body, rather than traversing through and connecting those mentioned only in the body, which is more time-consuming.

Two stocks may have LF relationships for various reasons. Firstly, news articles that mention multiple stocks tend to highlight leaders, such as dominant shareholders and customers, in the title, while placing followers, like suppliers and subsidiaries, in the main body. We refer to this type of LF link as an "inherent LF link", as it reflects a genuine directed relationship between two stocks. Besides this, there exists a second type of LF relationship, which is shaped by the way news is edited. This type of relationship does not necessarily reflect the genuine, direct leader-follower dynamics between stocks, but could be shaped by editorial interpretations and formatting styles. Previous research, such as studies by Chang and Suk (1998) and Dougal et al. (2012),

[1] This behavior aligns with the texton theory widely studied in psychology, as indicated in Julesz (1984). This theory suggests that without effort or scrutiny, differences in a few local conspicuous features (called textons) are detected over the entire visual field.

[2] See https://sherpapg.com/studies/attention-span-decline/.

has investigated how journalists' interpretations affect investor behavior. We term this type of LF link as a "format-induced LF link". In this paper, we provide new evidence that the news format independently exerts a significant impact on investors' information processing costs.

Given investors' limited attention and constrained information processing capacity, asset prices may respond to new information slowly. Higher information processing costs lead to slower information diffusion. A notable anomaly stemming from this limited attention and slow diffusion is "momentum spillover". This anomaly presents an ideal test case, as information propagates slowly through economic linkages.[3] Those linkages that are less perceptible to investors are likely to produce a more pronounced momentum spillover effect. Previous studies on momentum spillovers have typically tested the limited attention theory using firm-level characteristics (like size, analyst coverage, and trading volume) as proxies for a firm's received attention. However, our approach diverges from these studies. We propose that investors perceive different types of news-implied linkages differently due to their varying information structures and processing costs. This difference, we argue, naturally leads to momentum spillovers of differing intensities.

We examine the impact of information structure on processing costs and its subsequent effect on investor behavior and market outcomes. Central to our analysis is the assumption that news-implied linkages of varying complexity inherently attract differing levels of attention, as previously discussed. If processing leader-follower (LF) links is less costly, it should be easier to form trading signals based on leaders' performance. Consequently, this would result in faster information diffusion and a less pronounced cross-firm momentum spillover effect through LF linkages. Conversely, identifying peer (PE) links involves higher information processing costs, making the formation of trading signals more challenging. As a result, investors may underreact to related cross-firm signals, leading to a more significant cross-firm momentum spillover effect through PE linkages.

We conduct our empirical analysis using millions of Chinese news articles collected from the Financial Text Database of Financial Engineering Laboratory at Peking University. We focus on the Chinese stock market for three main reasons. Firstly, the Chinese stock market is dominated by retail investors, who generally have a more limited information processing capacity compared to institutional investors, as noted by Barber and Odean (2008). This is crucial for our underlying assumption to hold. Secondly, it has been shown that the 52-week high price effect is weak in China (Hou et al., 2023), allowing us to minimize the influence of psychological barriers among investors. Lastly, our test relies on the assumption that different news-implied linkages (LF, CM, and PE) involve different information processing costs, which in turn influence the intensity of momentum spillovers. Therefore, a key premise is the prominence of momentum spillover across these news-implied linkages. Ge et al. (2023) empirically demonstrates that news-implied momentum spillover is significant in the Chinese market, overshadowing other types of economic linkages. In contrast, from Scherbina and Schlusche (2013), such an anomaly is less pronounced in the U.S. market, where a trading strategy based on news-implied cross-predictability does not yield profits after accounting for trading costs. Therefore, the Chinese stock market presents an ideal setting for our study.

[3]The momentum spillover effect can take various forms because of the definitions of economic links. For example, industry links (Moskowitz and Grinblatt, 1999; Hou, 2007); customer-supply links (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010), alliance links (Cao et al., 2016), text-similarity links (Hoberg and Phillips, 2016, 2018), geographic links (Parsons et al., 2020; Jin and Li, 2020), technology links (Lee et al., 2019; Duan et al., 2022; Bekkerman et al., 2023), analyst co-coverage links (Ali and Hirshleifer, 2020), news-implied links (Scherbina and Schlusche, 2013; Guo et al., 2017; Schwenkler and Zheng, 2019; Ge et al., 2023; Wang, 2023), concept links (Du et al., 2022), and etc. The "momentum spillover" effect may also be called the cross-firm momentum or the cross-firm predictability. Throughout this paper, all these names refer to the same concept.

We first provide empirical evidence that the directed leader-follower (LF) links are easier to process than undirected peer (PE) links, leading to faster information spillover via LF links. Using the Baidu abnormal search volume index (ASVI) as the measurement of investors' attention (Da et al., 2011; Guo et al., 2017), we observe a quicker attention spillover through LF links compared to PE links. Specifically, for a focal firm, attention-grabbing events involving its leader stock result in a significant same-day increase in its ASVI by an average of 0.0102 (t-statistics = 5.15). In contrast, attention-grabbing events involving peer stocks show a notably lower and statistically insignificant impact with a coefficient of 0.0036 (t-statistics = 1.48) on the focal firm' s abnormal attention on the same day. However, past events involving peer stocks do predict a significant increase in abnormal attention to the focal firm, albeit with a delay. For example, the happening of attention-grabbing events from peer stocks in the past day lead to an increase of the focal firm's ASVI by 0.0071 (t-statistics = 2.51). This indicates that attention spillover via PE links occurs more slowly.

To assess the impact of information processing costs on investor behavior, we focus on momentum spillover as our test case. Using portfolio analysis and Fama and MacBeth (1973) regressions, we compare the performances of cross-firm momentum through LF, PE, and CM linkages.[4] If our story holds, the cross-firm momentum spillover effect should be more prominent via PE and CM links (undirected) and less prominent via LF links (directed). We begin by comparing portfolio sorting results, using signals computed from both directed and undirected linkages. At the end of each trading day, for each linkage type, we construct predictive signals for every stock based on the weighted average returns of its linked stocks.[5] With the predictive signals, we sort all sample stocks into quintiles and construct long-short portfolios by buying stocks from the top quintile and shorting stocks from the bottom quintile.

Overall, momentum spillovers via LF, PE, and CM links all yield significantly positive average returns and alphas. However, there are notable differences in the strength of these spillover effects. Specifically, the LF momentum spillover is substantially weaker compared to the PE and CM momentum spillover. To be precise, an equal-weighted long-short strategy based on LF momentum spillover yields an average return of 2.66% (t-statistics = 9.62) and a CH-4 adjusted alpha of 2.71% (t-statistics = 9.46). In contrast, an equal-weighted long-short strategy based on PE momentum spillover generates an average return of 5.11% (t-statistics = 12.91) and a CH-4 adjusted alpha of 5.14% (t-statistics = 12.97), which are double the figures observed in the LF case.

Furthermore, we examine the statistical significance of the return spreads between PE-based and LF-based strategies, and we find that this difference is both positive and statistically significant. In addition, the LF momentum spillover effect is also weaker than the CM momentum spillover effect, and the return spread between the CM-based and the LF-based strategies is positive and statistically significant. These results remain robust after controlling for the Fama and French (1993) three factors, Fama and French (2015) five factors, Carhart (1997) four factors, and Liu et al. (2019) Chinese four factors. The variation in investment gains clearly

[4] We include the cross-firm momentum driven by the original news co-mention link (CM) in our analysis to provide a comprehensive comparison, as CM represents an intermediate case between LF and PE. Additionally, since PE links are constructed by excluding all LF links from the news co-mention matrix, this approach aims to dispel any concerns that the separation is arbitrary. We seek to demonstrate that CM links are indeed distinct from LF links, which would, in turn, imply differences in the momentum spillover effect associated with each.

[5] For CM and PE links, the weight is the number of co-mentions times during the identification window; for LF links, the weight is the number of times the leader leads the follower during the identification window. Our main study is based on daily data, aiming to depict the dynamic of investors' reactions in a more timely fashion. As a robustness check, we have also conducted the exercise using weekly data.

illustrates how different information structures lead to different information processing costs and, ultimately, affect investment performance. The adage "no pain, no gain" holds true here; while identifying undirected linkages from the news is challenging, it yields better investment outcomes.

We then conduct Fama and MacBeth (1973) regressions to further investigate the differences in the predictive power of the LF, PE, and CM momentum spillover. In all regressions, the coefficient associated with the PE-based predictive variable consistently exceeds that of the LF-based variable. For example, on average, a one standard deviation increase in PE_Rtn predicts an increase of 3.13 bps (t-statistics = 6.93) in the future return, while a one standard deviation increase in LF_Rtn only predicts an increase of 0.76 bps (t-statistics = 2.32). The pattern is similar for the CM_Rtn. More importantly, after we control for PE_Rtn, both LF_Rtn and CM_Rtn lose their predictive power, with their coefficients dropping to an insignificant 0.0029% (t-statistics = 0.74) and 0.0035% (t-statistics = 0.52) respectively. Meanwhile, PE_Rtn maintains its significance with a coefficient of 0.0180% (t-statistics = 2.73). Additionally, in comparing the predictability over longer periods, spanning the next two to five trading days, we find that the PE momentum consistently shows the strongest predictability, while LF momentum diminishes the quickest among the three types of momentum spillover effects.

To further understand the nature of the LF relationship, we introduce a variation of the lead-follower linkage, termed LF2. This variant of the lead-follower linkage specifically excludes connections that do not exhibit a CM (co-mention) relationship, which we refer to as LF-only links, during our identification period. We posit that these LF-only links can effectively represent inherent leader-follower dynamics, exemplified by relationships between controlling shareholders (leaders) and their subsidiaries (followers). In Appendix B, we provide some examples of pairs that have only LF links. This analysis validates that the two stocks in an LF-only link exhibit distinct, nonequivalent positions, manifesting in scenarios such as supply-chain relationships with unequal status and in controller-subsidiary relationships. In contrast to LF-only links, LF2 links represent LF connections that arise from news format and are shaped by journalists' interpretations as we have removed the "inherent LF links". We develop a hypothesis to test whether news format can influence investors' information processing costs, consequently affecting market outcomes. Our empirical findings reveal that, compared with PE2 momentum,[6] LF2 momentum is significantly weaker, suggesting a notable impact of news formatting on market behavior.

One potential concern is that our result is driven by the findings in Hou (2007), specifically that the observed weak LF momentum spillover could be attributed to firm size. Hou (2007) found that information flows predominantly from large to small firms but not vice versa. Thus, the poor performance of the LF momentum may result from the blocked information transmission from small leaders to larger followers. To address this concern, we decompose the LF momentum into two parts according to the size of the leader firm: the big-leader LF momentum and the small-leader LF momentum. For the former, we only keep an LF link if the leader stock's market capitalization is in the top 50% of our sample. For the latter, we only keep an LF link if the leader stock's market capitalization is in the bottom 50% of our sample. If our results were entirely attributable to the size effect, we would expect to see no or weaker momentum spillover from small leaders. On the contrary, based on portfolio sorting analysis using equal-weighting, the momentum spillover from small leaders (small-leader LF momentum) is stronger than that from large leaders (big-leader LF momentum). When we apply value-weighting, we find no significant difference between the small-leader and big-leader LF momentum strate-

[6] During the identification window, we establish PE2 links by excluding all LF2 links from CM links.

gies. This indicates that the weaker LF momentum spillover observed in our study is not solely attributable to the size effect documented in Hou (2007). While Hou (2007) primarily focuses on the variation in information diffusion based on firm characteristics like size, our research takes a unique approach by concentrating on the characteristics of cross-firm linkages. This focus provides a distinct perspective on the phenomena of slow information diffusion and limited attention in the financial markets.

We performed a number of additional tests to ensure the robustness of our main conclusions. Firstly, we take transaction costs into consideration and re-compare the portfolio sorting results for three types of news-based momentum spillovers. After considering transaction costs, the long-short portfolio based on LF signals fails to generate statistically significant returns. In contrast, long-short portfolios based on PE signals and CM signals both continue to generate statistically significant positive monthly average returns and alphas. We also varied the identification windows (ranging from 30 to 180 days), employed different identification strategies for news co-mention (CM) links (including both *same_article* and *same_sentence* strategy from Ge et al. (2023)), and analyzed the data at different frequencies (from daily to weekly). Across all these specifications, our main finding that the cross-firm momentum spillover effect is less pronounced through LF links—remains robust.

Our paper offers three contributions. First, it demonstrates that processing information from the same source, in our context, a single news item, incurs different costs depending on the structure of the information. This difference in processing costs results in varying levels of attention being paid to different types of news-implied linkages. This contribution is closely related to a body of literature that has explored various measures of investor attention at the individual stock level, such as trading volume (Gervais et al., 2001; Barber and Odean, 2008; Hou et al., 2009), extreme daily returns (Barber and Odean, 2008), news and media coverage (Barber and Odean, 2008; Li and Yu, 2012; Chen et al., 2023), earnings announcements (Hirshleifer et al., 2011; Dellavigna and Pollet, 2009), analyst coverage (Hirshleifer and Teoh, 2003; Peng and Xiong, 2006; Hirshleifer et al., 2013; Ali and Hirshleifer, 2020), internet search volume (Da et al., 2011; Guo et al., 2017), and etc.[7] Different from these studies that focus on stock-level attention, we focus on attention paid to linkages. Limited research has been conducted on comparing investor attention with respect to firm linkages. Our paper innovatively constructs directed and undirected information structures based on news editing and readers' habits. This approach allows us to naturally categorize different types of news-implied linkages. This difference in information processing costs has important implications for the information transmission through these linkages and, consequently, the intensity of momentum spillover effects.

Secondly, our paper sheds new light on the mechanism behind the momentum spillover effect. Much of the existing literature attributes this anomaly to limited attention, with significant contributions from Cohen and Frazzini (2008), Cao et al. (2016), Lee et al. (2019), Ali and Hirshleifer (2020), Eisdorfer et al. (2022), Bekkerman et al. (2023), among others.[8] These studies typically employ indirect attention proxies for investor attention at the individual stock level, categorizing stocks into groups of high and low attention. They then analyze the heterogeneity in the strength of momentum spillover effects between these groups. Yet, this approach may overlook investors' attention to cross-firm linkages, which are inherently responsible for the existence of the spillover effect. After all, the fact that a stock attracts significant attention does not necessarily translate into

[7] See Chen et al. (2022) for a more comprehensive list of such studies.

[8] Apart from limited attention, some new psychology-based mechanisms have been recently proposed to explain the cross-firm momentum anomaly. For instance, Huang et al. (2021) attributed it to investors' psychological anchoring barrier to the 52-week high stock price; Huang et al. (2022) explained it in terms of information discreteness.

similar attention shifts for its linked stocks. In this paper, we argue that different information structures require different processing costs, leading to distinct levels of attention allocated to different types of linkages. This variation in attention has important implications for the intensity of the momentum spillover effect through these various linkages.

Thirdly, the paper relates to the literature that mines soft information from the news text. In recent years, there has been exploding empirical research in economics and finance that utilizes news texts to extract useful information. To mention a few, Hillert et al. (2014) rely on millions of news articles around the world and find that momentum predictability is strongest for firms in the spotlight of public attention; Calomiris and Mamaysky (2019) develop a classification methodology for news articles to predict risk and return in stock markets around the world; Obaid and Pukthuanthong (2022) propose a novel measurement of investor sentiment by analyzing news media pictures. More importantly, the interest in using news coverage to identify links among stocks is growing. Scherbina and Schlusche (2013) have found the effectiveness of stock connections identified through news co-mentioning in predicting stock returns. Additionally, Schwenkler and Zheng (2019) have developed a machine-learning method to construct a news-implied network of firms, and they also applied the same identification strategy to establish crypto linkages (Schwenkler and Zheng, 2021). Guo et al. (2017) have explored the relationship between news co-mentioning and investor attention spillovers. Ge et al. (2023) show that the same sentence co-mentioning strategy is more effective in identifying genuine links among firms than the traditional same article strategy. On their basis, this paper further distinguishes the news-implied link into the leader-follower (LF) link and the peer (PE) link according to the positions of stocks mentioned in the news. We propose that a directed leader-follower (LF) link comprising a leader firm that appears in the news title and a follower firm in the body is naturally more attention-grabbing, and thus the momentum spillover effect via such linkage should be weaker under the slow information diffusion and limited attention story.

The rest of the paper is organized as follows. In Section 2, we develop our main hypothesis. Section 3 introduces each type of link and describe the data. Section 4 provides evidence of the attention spillover through the three links. Section 5 compares the predictive power of the LF, PE, and CM momentum by portfolio sorting method and the Fama-Macbeth regression. Section 6 further examines the impact of news format on information processing costs. In Section 7, we conduct some further analysis and robustness checks. Section 8 concludes. Additional materials are given in Appendix.

2 Hypothesis Development

In this section, we explain the conceptual framework and develop our main hypothesis.

Different news-implied linkages possess varying information structures and therefore involve different information processing costs for investors to identify and utilize them in trading. We hypothesize that the LF (leader-follower) link, characterized by a more straightforward structure, is easier for readers to perceive. [9] On the contrary, the peer (PE) link between firms appearing only in the news body has a more complex information structure, making the linkage harder for readers to perceive.

We illustrate the intuition with a simple toy example.[10] As shown in Figure 1, consider four stocks named

[9]This prominence is partly driven by editors' efforts to create attention-grabbing headlines. Furthermore, due to the anchoring effect, as described by (Tversky and Kahneman, 1974), the first firm mentioned in the title often forms a psychological anchor, leaving a deeper impression on readers.

[10]We also provide two real news examples in Appendix B.

A to D: in Situation 1, stock A is the leader stock mentioned in the news title, while stocks B, C, and D are follower stocks only appear in the news body. Furthermore, in this case, the information structure is so clear that investors only need to process information from stock A (there are only three directional LF links in total). In contrast, in Situation 2, where all four stocks are mentioned solely in the news body and are peer stocks with each other, the information network is significantly more complex, encompassing 12 undirected PE links. Additionally, since the four stocks are presented with equal prominence, it becomes more time-consuming for investors to traverse through all four stocks to detect shocks and formulate corresponding peer-based trading strategies.

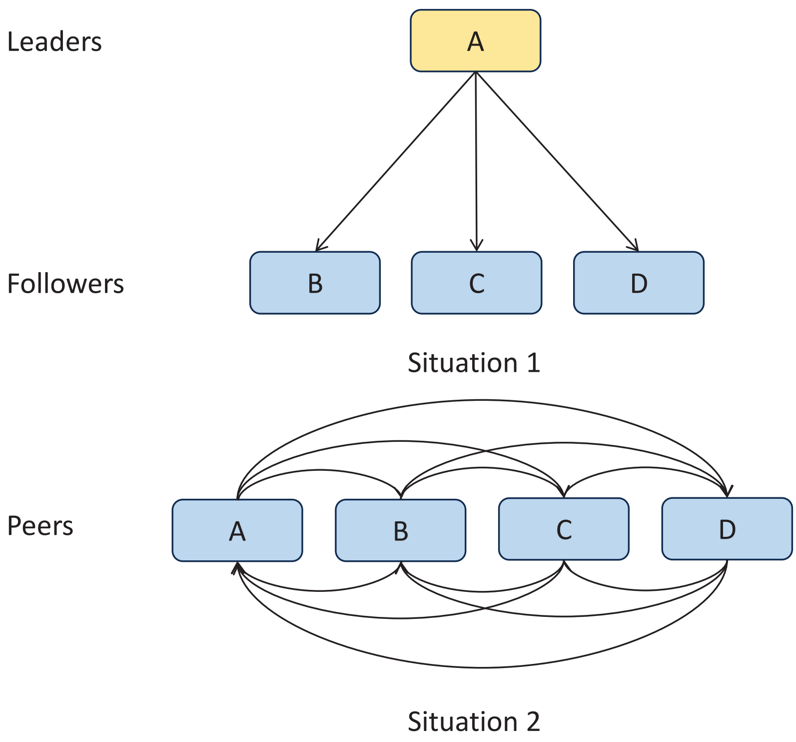

Figure 1: An illustration of the news information structure

According to Guo et al. (2017), a news article mentioning multiple stocks creates an attention spillover effect across these stocks. Therefore, if a LF link is easier for investors to identify compared to a PE link, we would expect the spillover of investor attention to be faster through the LF link. Still using Figure 1 for illustration: in Situation 1, there's an LF link from leader stock A to follower stock B and a PE link between stock B and C. When stock A experiences an "attention-grabbing" event, increasing investor attention towards it, stock B should also attract attention more swiftly, given the easily recognizable LF link. Conversely, with the PE link, an investor would need to analyze all stocks from B to D to determine whether any peer stock experiences a shock, a process that is inherently more time-consuming and costly. For instance, if a shock occurs to stock C, the attention would then spill over to stock B, but with a delay. In this scenario, the speed of attention spillover from stock A to stock B is faster than from stock C to stock B. Therefore, we summarize hypothesis 1 below:

Hypothesis1. *The information processing cost associated with LF links is lower than that of PE links. When leader firms experience attention-grabbing events, investors can rapidly shift their focus to the focal firm due to the straightforward information structure of LF links. Conversely, when peer firms experience shocks, there is a slower spillover of investor attention towards the focal firm, attributed to the more complex information* *structure inherent in PE links.*

We now explore the impacts of information processing costs on investors' behaviors and market outcomes by examining momentum spillover effects via different types of news-implied linkages. In an efficient market, a firm's stock price should quickly adjust to news from its economically connected firms as soon as that information becomes public. However, in reality, due to limited attention and information processing capabilities, investors may not immediately incorporate relevant information, resulting in slower propagation of information through economic linkages. Consequently, the speed of information diffusion is likely to be inversely related to the information processing costs. In summary, within the framework of limited attention, the cross-firm predictability of linked firms to a focal firm is often a result of investors' inattention to the linkage. Thus, we argue that if a link between firms is more easily noticed by investors, the associated momentum spillover effect driven by this link should be weaker.

In Situation 1 of Figure 1, when stock A experiences a shock, investors will quickly react to stocks B, C, and D due to the direct LF relationship. As a result, the momentum spillover driven by the LF link from stock A to B, C, and D is expected to be weaker and dissipate more swiftly. Conversely, in Situation 2, where the PE network involves four stocks of equal status in an undirected network, an investor must first identify which stock has experienced a shock and then evaluate the remaining three stocks for investment opportunities. This process leads to stronger and more persistent lead-lag predictability in momentum spillover via PE links. Accordingly, we state the following hypothesis:

Hypothesis 2. *A lower information cost is associated with a weaker momentum spillover effect and a faster decay in predictability. Consequently, the momentum spillover effect via LF linkages is weaker and dissipates more quickly than via PE linkages.*

Admittedly, media articles provide investors with valuable "public information". On the one hand, journalists disclose facts about events, including relevant stocks. On the other hand, they can infuse their interpretations into news reports, as suggested by Dougal et al. (2012) and Blankespoor et al. (2020), such as using attention-grabbing titles. This motivates us to distinguish the effects caused by the "inherent LF link", which arise from genuine directed relationships, from those primarily due to journalistic writing styles, i.e., the "format-induced LF link".

It is noteworthy that if a pair of stocks, say A and B in Situation 1 (as shown in Figure 1), appear only in an LF relationship (termed LF-only links), there is an inherent asymmetry. In such instances, news articles that co-mention A and B should consistently feature stock A in the title during the identification window. We show examples of LF-only links in Appendix B, which validates that the two stocks in an LF-only link exhibit genuine leader-follower relationships, manifesting in scenarios such as supply-chain relationships with unequal status and in controller-subsidiary relationships.

In this paper, we focus on the effect of news format on attention and information costs. We assume that inherent LF relationships are well approximated by LF-only links. Then, we create LF2 links, which exclude LF-only connections from LF links, to represent LF relationships induced by news format. LF2 links indicate that stocks A and B have both an LF link, as in Situation 1, and a PE link, as in Situation 2, during the identification window. Consequently, LF2 links represent LF connections that arise from news format and are shaped by journalists' interpretations as we have removed the "inherent LF links".

If the format of news does affect information processing costs, we would expect information to diffuse more rapidly along LF2 linkages compared to PE2 linkages. Consequently, the momentum spillover effect through

LF2 links should be weaker than that through PE2 links. We summarise our hypothesis as follows:

Hypothesis 3. *The formatting of news reports influences information processing costs. Reports featuring leading stocks in the title can capture investors' attention, making it easier for them to recognize follower stocks. Consequently, this results in a weaker and less persistent momentum spillover effect.*

3 Data

The sample stocks in this paper contain all A-shares listed on the main boards of the Shanghai Stock Exchange (SSE), Shenzhen Stock Exchange (SZSE), and the Growth Enterprise Market (GEM). Special treatment shares are excluded. The sample period starts from January 2012 to December 2021 for two reasons: firstly, it is consistent with Ge et al. (2023) to guarantee the quality and quantity of news; secondly, the Baidu search volume index (SVI) used to study the investors' attention spillover has only been applicable since 2011, and a lag of one year is needed to construct the abnormal search volume index (ASVI).

In later subsections, we provide detailed information about the identification methodologies for different stock relationships inferred from news including the news co-mention link (CM), leader-follower link (LF), and peer link (PE). Then we describe the data sources and constructions of other key variables used in the paper, followed by the summary statistics.

3.1 Link Identifications

The identifications of the three kinds of news-implied links rely on textual analysis of news about the Chinese stock market. We use millions of news articles from the Financial Text Database of Financial Engineering Laboratory at Peking University [11] starting from 2006 to 2020. We screened out 1,138,247 news articles that mentioned at least one listed firm in the A market. Table 21 from Appendix A reports the summary of the daily basic information of the news data since 2006. Prior to 2012, the news data was relatively sparse, with an average of fewer than 100 news pieces per day. The news data has become much more abundant since 2012. Given that, although the news data has been available since 2006, in the main body of the paper, we use the subset from 2012 to 2021 when the news data quality is high.

3.1.1 News Co-mention (CM)

We first introduce the news co-mention link (CM) which is the foundation of the further identifications of the leader-follower link (LF) and peer link (PE). News co-mention has been widely used as a way to define firm links (see e.g. Guo et al. (2017); Schwenkler and Zheng (2019); Ge et al. (2022); Wang (2023)). The above literature defines two firms to be connected if they are co-mentioned in the same news article. However, Ge et al. (2023) has found that, in the context of China, the *same_sentence* strategy that defines two firms to be connected only if they are co-mentioned in the same sentence of the news could preclude potential noisy connections defined by the *same_article* strategy. Thus, the *same_sentence* strategy has a superior performance than the ordinary *same_article* strategy. Moreover, the cross-firm momentum driven by the *same_sentence* type of co-mention shows the unifying effect over other kinds of momentum spillover effects. Consistent with their study, we define

[11] See: https://finlab.pku.edu.cn

two firms to have a CM link if they are co-mentioned in any of the same sentences in a news article during the past 90 days.[12]

At the end of each trading day, for a focal firm i, its CM predictive signal, i.e. CM_Rtn_i, is computed as the average return of all its CM linked firms weighted by the number of co-mention times during the identification window.

3.1.2 Leader-follower (LF)

In a piece of news, the title is usually the most conspicuous because it is often edited so deliberately by the author or journalists to catch the eyes of potential readers (Hu and Härdle, 2021). Meanwhile, due to people's reading habits and limited information processing capacity, they are easier to give more attention and importance to the news title. Therefore, stock names or codes appearing in the title should be more special than those in the news context. During the pre-specified identification window, we define the stock shown in the news title as the leader stock, because it is often selected by the writer on purpose to represent the related group of stocks to the news; besides, as title stocks, they also catch investors' attention much more easily. Accordingly, stocks appearing in the body part of the *same* news are referred to as the followers of the leader stock. The link from the leader stock to the follower stock is the leader-follower link (LF).

Unlike the news co-mention link which is undirectional, the leader-follower link is directional in which the information flows from the leader stock to the follower stock.[13] During the identification window, the same LF pair may be identified by multiple news, and we take the number of corresponding news as the weight to measure the strength of the LF relationship. At the end of each trading day, for a focal firm i, its LF predictive signal, i.e. LF_Rtn_i, is computed as the average return of all its leader firms weighted by the number of the leading times during the identification window.

3.1.3 Peer Effect (PE)

During the identification window, some firms have only been identified as CM-connected firms and the LF relationship has never formed. Specifically, these firms are co-mentioned only in the body part of news articles and none of them have appeared in the title. We call these firms the peer firms and the relationship between these firms as the peer relationship (PE). From the definition, PE links could be constructed based on the CM links and LF links of the same identification window. What we need to do is subtract all LF links from CM links during the identification window. The peer relationship matrix is given by the following formula:

$$\mathbf{PE}_t = \mathbf{CM}_t \odot \mathbf{I}[\mathbf{LF^*}_t]$$

$$\mathbf{LF^*}_t = \mathbf{LF}_t + \mathbf{LF}_t^{'}$$

where $\mathbf{PE}_t$ is the peer matrix at time t, $\mathbf{CM_t}$ is the news co-mention matrix at time t, $\mathbf{LF_t}$ is the leader-follower matrix at time t. The operator $\odot$ means taking the product of the corresponding elements of two matrices of the same dimension. $\mathbf{I}[\cdot]$ is the function that is applied to each element of the given matrix, and returns 1 if the element value is 0, otherwise returns 0. [14]

[12] We report the results under other identification windows in the robustness part.

[13] For example, if stock A is mentioned in the titles of three different news articles in the past 90 days, and meanwhile stock B is mentioned in the body of each article, then we say that stock A leads B three times.

[14] In Appendix C we give a simple case to illustrate the construction of the **PE** matrix

Our identification method assumes that the leader-follower relationship has a higher identification priority than the CM link and PE link for the same pair of stocks. That is, even if the two firms have been identified to be news co-mention linked firms or peer firms several times, as long as there has been one LF link formed between them, we finally only recognize them as LF connected firms. This treatment is meaningful for two reasons: firstly, according to the analysis in the development of **Hypothesis 1**, due to investors' limited information processing capacity and reading habits, they tend to pay more attention to LF link than PE link; secondly, it is not likely for investors to "forget" the LF link during the short identification window. As a result, once the LF link is formed, the CM link or PE link identified before or after will be replaced, but the vice versa is not true.[15]

At the end of each trading day, for a focal firm i, its PE predictive signal, i.e. PE_Rtn_i, is computed as the average return of all its peer firms weighted by the number of the co-mention times during the identification window.

3.2 Other Key Variables

Our paper includes a series of other variables. Specifically, the daily return (Rtn) is the daily rate of return taking into account the reinvestment of dividends. The daily turnover rate ($Turnover$) is computed as the ratio of the daily trading volume and the number of outstanding shares. The illiquidity indicator ($Illiquidity$) is computed according to Amihud (2002). The market beta ($Beta$) is estimated using daily returns in the past 120 trading days, and the market return is the composite return of LF-connected Shenzhen markets. The return on equity (ROE) is computed as the net income divided by the stockholders' equity. The book-to-market ratio (BM) is calculated as the ratio of the shareholders' equity over the market capitalization. The earnings-to-price ratio (EP) is calculated as the ratio of the net income per share over the stock price. The total assets growth rate ($Assets\_growth$) is calculated as the percent change in the total assets for the current year to the last year.

To evaluate retail investors' attention on one specific stock, we construct the abnormal search volume index ($ASVI$) following Da et al. (2011) and Guo et al. (2017). The daily abnormal search index ($ASVI$) is computed as the percentage change between the daily search volume index (SVI) for a stock and its past 1-year mean (Guo et al., 2017), skipping the most recent day. However, there are two slight differences from the US study. Firstly, since we focus on China's A-share market, the search engine we use is Baidu rather than Google.[16] Secondly, the search keywords for US stocks are only the tickers of stocks (like AAPL for Apple), while in China, our search keywords not only include the stock code but also the Chinese abbreviation name and the Chinese full name of the firm. This is because, in China, each stock is given a unique 6-digit code which is hard for most investors to remember. In most situations, investors prefer searching the Chinese abbreviation name rather than the digit code of the firm.[17] Moreover, even if some investors prefer searching the stock

[15] So subtracting the $\mathbf{CM}_t$ and $\mathbf{LF}_t$ directly, i.e., $\mathbf{PE}_t = \mathbf{CM}_t - \mathbf{LF}_t$, is meaningless.

[16] Globally, Baidu is the third largest search engine only to Google and Bing, but in China, Baidu is undoubtedly the most popular search engine. It is also the largest Chinese search engine in the world. According to Baidu's 2021 annual report, Baidu's monthly active users reached 622 million.

[17] Take Contemporary Amperex Technology Co. Limited (CATL) whose stock code is 300750 as an example. CATL is the leading manufacturer of new energy batteries around the world and also the hottest stock in the new energy concept. However, the daily search volume of its stock code (300750) is so low that the Baidu Index does not even include the code as a search keyword. On the contrary, its Chinese abbreviation name (宁德时代) experienced a very high search volume during the past few years, peaking

code directly, they need to search the firm name first to get the digit code. As a result, when constructing the search volume index of Chinese stocks, it is better to include firms' Chinese names in search keywords. We use $ASVI\_Code$ and $ASVI\_All$ to respectively indicate the abnormal search volume index based on codes only and with all codes and names.[18]

Our data come from various sources. Stock trading data and financial statement data are available from CSMAR. The Baidu search volume index (SVI) is only available after 2011 and could be directly obtained from the Chinese Research Data Services Platform (CNRDS).[19] In the portfolio sorting analysis, the factor data of Fama and French (1993) 3-factor model and Fama and French (2015) 5-factor model are provided by CSMAR. The daily Carhart (1997) 4-factor data is collected from BetaPlus.[20] The daily CH-4 factor data is obtained from Prof. Robert F. Stambaugh' s website.

3.3 Descriptive Statistics

Table 1 shows the descriptive statistics. We divide all sample stocks into three groups and make summary statistics respectively. At the end of each trading day, if one firm has only been identified as the leader stock during the past 90 days, it is included in the leader group. The follower group and peer group are defined in the same way. Panels A, B, and C show the summary statistics of leader, follower, and peer stocks respectively, while Panel D shows the global statistics of the full sample.

On average, the abnormal search volume index based on codes only ($ASVI\_Code$) and both codes and names ($ASVI\_All$) of leader stocks are 0.30 and 0.31, which are higher than those of both follower stocks (0.24 and 0.23) and peer stocks (0.26 and 0.27). This finding verifies our expectation that stocks appearing in the title could trigger more attention from investors than stocks mentioned in the news content. More importantly, these differences are statistically significant. In Table 2, we report the results of the two-sample t-test of these characteristics among leaders, followers, and peers. The $ASVI\_Code$ and $ASVI\_All$ of leader stocks are 5.80% and 8.15% higher than follower stocks, with t-statistics at 16.50 and 20.79 respectively. Meanwhile, the two attention proxies of leader stocks also outperform those of peer stocks by 4.05% (t-statistics = 12.68) and 3.97% (t-statistics = 9.59) respectively.

4 Speed of Attention Spillover

As is proposed in **Hypothesis 1**, if investors are indeed more likely to pay attention to the LF link rather than the PE link, then we can expect that the spillover speed of investor attention should also be faster through the LF link. Specifically, when there is an "attention-grabbing" event happening to the leader stock, the investor's attention on the follower stock should increase swiftly. However, the investor attention shifts among peer stocks should be more sluggish.

We use the Baidu abnormal search volume index to measure the investors' attention to one specific stock to test **Hypothesis 1**. As documented by Da et al. (2011), when an investor searches a stock keyword on the search engine, he/she must be paying attention to the stock. Therefore, the stock search frequency on the search

at more than 30,000 searches a day in 2022.

[18]We report the summary statistics for both $ASVI\_Code$ and $ASVI\_All$, while in empirical analysis, we rely on $ASVI\_All$ for identification accuracy.

[19]Baidu does not provide access to earlier search data before 2011.

[20]See: https://www.factorwar.com/data/factor-models/

engine is a more straightforward measurement of attention than indirect proxies like stock abnormal returns, trading volume, and news. The detailed construction process of the Baidu $ASVI$ is given in subsection 3.2.

Furthermore, we define two dummy variables based on $ASVI$ values of leaders and peers of each stock: $LF\_Events_{i,t}$ and $PE\_Events_{i,t}$. These two dummy variables capture whether the leader stocks and peer stocks of one stock have attention-grabbing events on the trading day, respectively. For $LF\_Events_{i,t}$, if the $ASVI$ of any of the leaders of stock i ranks in the top 10% of all samples on trading day t, the dummy variable $LF\_Events_{i,t}$ takes 1, otherwise, the dummy variable takes 0. Similarly, the dummy variable $PE\_Events_{i,t}$ takes 1 if the $ASVI$ of any of the peers of stock i on the trading day t is in the top 10% of all samples, and 0 otherwise.

To examine whether the attention spillover is faster through the LF link than the PE link, we set up two fixed-effect panel models:

$$ASVI_{i,t} = FE + LF\_Events_{i,t+n} + PE\_Events_{i,t+n} + \mathbf{Control}_{i,t}, n = -10, -9, ..., 0, ..., 9, 10$$

$$ASVI_{i,t} = FE + \sum_{n=-10}^{10} LF\_Events_{i,t+n} + \sum_{n=-10}^{10} PE\_Events_{i,t+n} + \mathbf{Control}_{i,t}$$

In the first regression model, the dependent variable $ASVI_{i,t}$ is the Baidu abnormal search volume index of stock i at time t; FE indicates the entity and time fixed effect; the core explanatory variables are the two dummies $LF\_Events_{i,t+n}$ and $PE\_Events_{i,t+n}$, where n is an integer ranging from -10 to 10. If $n < 0$, this regression measures the impact of the *past* attention-grabbing events of leaders and peers on investor attention of stock i; if $n > 0$, this regression measures the impact of the *future* attention-grabbing events of leaders and peers on investor attention of stock i; if $n = 0$, this regression measures the impact of the attention-grabbing events of leaders and peers on investor attention of stock i at *the same* trading day. Therefore, with different values of n, this model results in a total of 21 regressions. The control variables include the daily return, the stock size (taking the logarithms), and the daily turnover rate. In the second panel regression, we add all the dummy variables $LF\_Events_{i,t+n}$ and $PE\_Events_{i,t+n}$ under different n to the regression. The setting of other variables is consistent with the first model.

Panel A and Panel B of Table 3 report the results of 21 regressions under the first model and 1 regression under the second model respectively. Standard errors are clustered by stocks and time. The values in parentheses give t-statistics. In Panel A, each row shows the result of the regression with the given n. Row $n = 0$ shows the impact of attention-grabbing events of leaders and peers on the investor attention of the stock i on the same day. The coefficient of LF_Event illustrates if a leader of stock i has an attention-grabbing event on that day, the $ASVI$ of stock i will increase by 0.0394 on average, which is higher than the impact of PE_Event of 0.0283. This result shows that in the short term, the speed of attention spillover of investors is indeed faster through the LF link than in the PE link. However, with the decrease of n from 0 to -10, the coefficient of LF_Event drops faster than that of PE_Event. To be specific, for regressions with n less than -7, the coefficient of LF_Event becomes lower than that of PE_Event. This finding further validates our conjecture that the attention spillover is faster in the LF link and its impact will be quickly assimilated, therefore, the impact of past attention-grabbing events of leaders on stock attention is lower than that of peers.

Panel B of Table 3 shows the result of the second regression model where all attention-grabbing event dummy variables are added and the conclusion is clearer. On the same day ($n = 0$), attention-grabbing events of the leader stocks will increase the investor attention of the focus stock significantly by 0.0102 (t-statistics = 5.15) on average. However, the happening of attention-grabbing events of peer stocks have a much lower and insignificant impact on investor attention, with an average increase of 0.0036 (t-statistics = 1.48). Again, these results show that investor attention is easier to spread through the LF link, which indicates that investors are more likely to notice the LF link. However, they could not recognize the PE link in time, as the attention could not spread from peer stocks to the focal stock on the same day.

On the contrary, there is a delay of the attention spillover through the PE link, since the impact of the attention-grabbing events of peer firms occurring one day ago on the focal stock attention is larger and more significant than the impact of the event occurring on the same day. As shown in row $n = -1$, the happening of attention-grabbing events of peers on the previous day will lead to an increase of 0.0071 (t-statistics = 2.51) in the $ASVI$ of the focal stock on the current day, while the past-one-day attention-grabbing events of leaders only lead to an increase of 0.0068 (t-statistics = 3.58) in the $ASVI$ of the focal stock on the current day. This is because the attention-grabbing events of the leader stocks on the previous day are quickly digested through the LF link on the same day, while the attention-grabbing events of the peer stocks on the previous day need to lag one day before being noticed by investors through the PE link.

In all, the regression analysis in this part provides direct evidence that the spillover speed of attention-grabbing events is faster through the LF link than through the PE link, thus investors are more likely to notice the LF link rather than the PE link.

5 Return Predictability

In this section, we will verify **Hypothesis 2** by testing and comparing the predictability of the cross-firm momentum driven by the leader-follower link (LF) and peer link (PE). Since the LF link is easier to notice by investors than the PE link and the CM link, the speed of information diffusion should also be faster through the LF link, which will lead to weakened underreaction of follower stock prices to the news of leader stocks (Hou et al., 2009). Therefore, since the news-driven cross-firm momentum is caused by limited investor attention, we should observe that the LF momentum underperforms the PE momentum and CM momentum significantly.

We first conduct the portfolio sorting analysis for each cross-firm momentum strategy. Then we control for more variables by the Fama and MacBeth (1973) regression and compare the coefficients of different momentum spillover signals. We also compare the predictive power of the three momentum over longer horizons. Finally, we conduct a placebo test by generating LF and PE links randomly from CM links.

5.1 Portfolio Sorting Analysis

For the cross-firm momentum based on the LF link (referred to as the LF momentum later), at the end of each trading day, all sample stocks are sorted into quintiles according to their LF signals: LF_Rtn. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short strategy is buying the highest signal group and selling the lowest signal group. All portfolios are held for one day and are rebalanced daily. In addition to reporting the average return, we also report the alpha of each portfolio using the Fama and French (1993) three-factor model (FF-3), Fama and French (2015) five-factor model (FF-5), Carhart (1997) four-factor model (Carhart-4), and Liu et al. (2019) Chinese four-factor model (CH-4). The daily returns and alphas are all converted to a monthly frequency for better comparability with existing literature. The PE momentum and CM momentum strategies are constructed similarly by replacing the sorting signal variable LF_Rtn with PE_Rtn and CM_Rtn respectively.

Table 4 presents the performances of the three momentum spillover strategies based on the identification window length of 90 days.[21] In general, all of LF, PE, and CM momentum could achieve significant and positive long-short mean returns and alphas. All Spearman rank correlation coefficients equal to 1 imply that portfolio returns or alphas are monotonically increasing with past momentum signals.

However, there are obvious differences in long-only and long-short performances of the three momentum strategies. Overall, the LF momentum performs the worst, while the PE momentum performs much better than the other two. Specifically, for equal-weighted portfolios, the LF momentum only generates a long-short mean return and CH-4 adjusted alpha at 2.66% (t-statistics = 9.62) and 2.71% (t-statistics = 9.46) per month. However, the PE momentum yields almost twice as much as the LF momentum, with a long-short mean return and CH-4 alpha at 5.11% (t-statistics = 12.91) and 5.14% (t-statistics = 12.97). Consistently, for the long-only strategy (the fifth portfolio), the LF momentum only generates a mean return and CH-4 alpha at 3.69% (t-statistics = 4.50) and 3.30% (t-statistics = 4.66) per month, while the PE momentum generates a mean return and CH-4 alpha at 5.34% (t-statistics = 6.33) and 5.29% (t-statistics = 6.61) per month. This return spread still exists within value-weighted portfolios or using other factor models including the FF-3, FF-5, and Carhart-4.

Furthermore, we can also observe a better performance of the PE momentum than the CM momentum. As shown in Panel C of Table 4, the CM momentum yields an equal-weighted long-short mean return and CH-4 adjusted alpha at 4.26% (t-statistics = 13.36) and 4.27% (t-statistics = 13.39), lower than those of the PE momentum at 5.11% (t-statistics = 12.91) and 5.14% (t-statistics = 12.97) respectively. Besides, the long-only PE momentum strategy also beats the long-only CM momentum. Considering that the PE link originates from the CM link, the better performance of the PE momentum than the CM momentum indicates that the PE link is not a homogeneous portion of the CM link and thus our decomposition of the CM link is both economically and statistically meaningful.

In Table 5, we further check the statistical significance of the return spreads by taking time series differences among the long-only and long-short portfolio returns of the three momentum strategies. As indicated in Panel A, for equal-weighted portfolios, the PE momentum outperforms both the LF momentum and CM momentum significantly. On average, the PE momentum achieves a monthly long-only mean return of 1.60% (t-statistics = 8.03) and 0.44% (t-statistics = 3.99) higher than the LF momentum and the CM momentum respectively. The differences in long-short returns are more significant, with 2.39% (t-statistics = 9.08) and 0.83% (t-statistics = 5.94) per month of the PE momentum higher than the LF and CM momentum respectively.

The portfolio sorting and comparing analysis demonstrate that the cross-firm momentum driven by the PE link is stronger than by the LF link or by the CM link both economically and statistically. This finding verifies our second hypothesis that investors' information cost on the LF link is lower than on the PE link.

5.2 Regression Analysis

We then further compare the predictability of the LF, PE, and CM momentum by Fama and MacBeth (1973) regressions. The advantage of using the regression analysis is that we could control for all three momentum

[21] Our conclusions are robust to the length of the identification window. We also report the results of portfolio sorting and Fama-Macbeth regressions under shorter (30 days) and longer (180 days) identification windows in the robustness part.

signals as well as a series of other variables at the same time. The basic regression model is set as below:

$$Rtn_{i,t+1} = 1 + \beta_{LF} LF\_Rtn_{i,t} + \beta_{PE} PE\_Rtn_{i,t} + \beta_{CM} CM\_Rtn_{i,t} + \boldsymbol{\beta}_{ctrl}^{'} \mathbf{Control}_{i,t}$$

where $Rtn_{i,t+1}$ is the daily stock return of firm i on the next trading day; $LF\_Rtn_{i,t}$, $PE\_Rtn_{i,t}$, and $CM\_Rtn_{i,t}$ are the LF, PE, and CM momentum predictive signals of firm i at time t, as constructed in subsection 3.1; $\mathbf{Control}_{i,t}$ is the control variable vector of firm i at time t and includes the firm size ($Size$, taking logarithms), book-to-market ratio (BM), earnings-to-price ratio (EP), return on equity (ROE), past daily return (Rtn), total assets growth rate ($Assets\_growth$), and daily turnover rate ($Turnover$). Non-return variables are winsorized at 1% and 99% and all independent variables are cross-sectionally standardized.

Table 6 reports the results of Fama-Macbeth regressions. All coefficients are shown in percent. In columns 1 - 3, we add only one of the three momentum spillover signals (LF_Rtn, PE_Rtn, CM_Rtn) respectively to regressions. All coefficients of the three predictive variables are significantly positive, however, their economic magnitudes differ a lot. The coefficient of CM_Rtn is smaller than that of PE_Rtn, and the coefficient of LF_Rtn is further smaller than that of PE_Rtn. Specifically, on average, a one standard deviation increase in LF_Rtn and CM_Rtn only predicts an increase of 0.76 bps (t-statistics = 2.32) and 2.40 bps (t-statistics = 6.83) in the future return respectively, while for the PE momentum, a one standard deviation increase in PE_Rtn predicts an increase of 3.13 bps (t-statistics = 6.93) in the future return.

Furthermore, in column 4, we control for LF_Rtn and PE_Rtn simultaneously. The coefficient of PE_Rtn is about five times higher than that of LF_Rtn. More importantly, the coefficient of LF_Rtn becomes insignificant after controlling for PE_Rtn. On average, a one standard deviation increase in LF_Rtn only predicts a trivial increase of 0.40 bps (t-statistics = 1.09) in the future return, while for the PE momentum, a one standard deviation increase in PE_Rtn predicts an increase of 2.14 bps (t-statistics = 4.02) in the future return. Similarly, in column 5, after controlling for the PE_Rtn, the coefficient of CM_Rtn becomes insignificant with a t-statistics of 1.78.

Finally, in the last column of Table 6, we regress the future return on all three predictive signals. The coefficient of LF_Rtn decreases to only 0.29 bps with an insignificant t-statistics of only 0.74. Besides, the coefficient of CM_Rtn also becomes insignificant at 0.35 bps (t-statistics = 0.52). In contrast, the PE momentum still keeps its strong predictive power, and a one standard deviation increase in PE_Rtn predicts an increase of 1.80 bps (t-statistics = 2.73) in the future return. Ge et al. (2023) has found the momentum spillover effect driven by the news co-mention link (the CM link) has a unifying effect over other economic links (like the industry link, technology link, shared-analyst link, etc.), and after controlling the CM predictive return, other connected returns become insignificant; however, in our analysis, the PE momentum has even stronger predictive power and could even subsume the predictive power of the CM momentum.

In conclusion, our results of Fama-Macbeth regressions are consistent with the portfolio sorting analysis that the predictive power of the PE momentum is much stronger than the LF and CM momentum, which is also in line with our hypothesis that the PE link is harder for investors to recognize than the LF and CM link and thus the lead-lag effect of the PE momentum is more pronounced.

5.3 Predictability over Longer Horizons

In this subsection, we further compare the persistence of the predictive power of the LF, PE, and CM momentum over longer horizons. We follow Guo et al. (2017)'s method to use the stock returns over the next two to five days rather than the next one day as the dependent variable and re-conduct the Fama-Macbeth regression.

Table 7 reports the regression results. Overall, the coefficients for all three types of momentum weaken as the forecast gap increases. However, consistent with our previous conclusion, the PE momentum always outperforms the LF momentum a lot. For instance, a one standard deviation increase in PE_Rtn predicts an increase of 2.55 bps (t-statistics = 5.79) in the next fifth day's return, which is more than two times higher than 1.00 bps (t-statistics = 3.30) of LF_Rtn.

More importantly, the predictive power of the LF momentum also decays faster than that of PE momentum and CM momentum. The coefficient of LF_Rtn on the future return at $t + 2$ is 0.017% (t-statistics = 5.33), while the coefficient on the future return at $t + 5$ decreases about 41% to only 0.01% (t-statistics = 4.53). In contrast, from Rtn (t+2) to Rtn (t+ 5), the coefficients of PE_Rtn and CM_Rtn decrease by only 28% and 34% respectively.

The longer-horizon analysis further validates **Hypothesis 2** that, the LF momentum, which attracts more investors' attention, performs consistently weaker than the PE momentum, and its predictive power also weakens faster, which further proves that the information processing cost of the LF links is expected to be lower than that of the PE link.

5.4 Placebo Tests

In this part, we conduct placebo tests by generating the LF link randomly from the CM link rather than by identifying leaders and followers. At the end of each trading day, we randomly define 50% of CM links as LF links (denoted as Placebo LF), then the other 50% as PE links (denoted as Placebo PE). We re-conduct the portfolio sorting analysis based on predictive signals computed from the random connection metrics and compare their differences.

Table 8 reports the single sorting results of the placebo LF and PE momentum. In this setting, there is no clear difference between the two kinds of momentum. The equal-weighted placebo LF momentum generates a long-short mean return and CH-4 alpha of 4.09% (t-statistics = 12.07) and 4.11% (t-statistics = 12.14) which is only slightly higher than the placebo PE momentum of 3.94% (t-statistics = 10.86) and 3.96% (t-statistics = 10.89) respectively. The difference is also not statistically significant as shown in Table 9, with a long-short difference of -0.12% (t-statistics = -0.92) and a long-only difference of -0.15% (t-statistics = -0.80). More importantly, neither of the Placebo LF and PE momentum beat the CM momentum. For value-weighted long-short differences in Panel B of Table 9, the CM momentum even beats the Placebo PE and LF momentum significantly with a difference of 0.83% (t-statistics = 3.25) and 0.56% (t-statistics = 2.98) per month.

In general, the two momentum spillover effects driven by the LF and PE links randomly generated from the CM links do not differ significantly, and both perform worse than CM momentum. This result shows that identifying the LF link through the title-body relationship in the news and then obtaining the PE link by decomposing the CM link is not a random behavior, but does have economic significance.

6 The Impact of News Format

In this section, we further study the impact of the news format on investors' attention and information costs. We distinguish the LF link into two types. The first type is the "inherent LF link" that arises from genuine directed relationships. The inherent LF link is formed endogenously between firms that have real affiliation or dependence relationships. For such firms, if they appear in the news, they can basically only show LF links. On the contrary, the second type is the "format-induced LF link" that is formed by journalistic writing styles. Such connected firms may not have obvious endogenous affiliations or dependencies but simply are edited by journalists to have LF links.

We posit that if one LF link has never been identified as any CM link during the identification window, then it is the inherent LF link. We call it the LF-only link for brevity.[22] By removing all LF-only links, format-induced LF links could be obtained which are denoted as LF2 links. Accordingly, PE2 links are constructed by subtracting all LF2 links from CM links. If the format of news does have an impact on information processing costs, we would expect that information should diffuse more rapidly through LF2 links than PE2 links. Therefore, the cross-firm momentum driven by LF2 links should be weaker than that by PE2 links.

Table 12 reports the portfolio sorting performance of the LF2 and PE2 momentum, and Table 13 reports the portfolio return spreads between LF2, PE2, and CM momentum. Consistent with our previous analysis, the LF2 momentum is significantly weaker than the PE2 and CM momentum. Table 14 further provides with the Fama-MacBeth regression results. When predictors are added to the regression alone, their coefficients are all significant, specifically, PE2 > CM > LF2. However, after controlling for $PE2\_Rtn$, both $LF2\_Rtn$ and CM_Rtn lose their predictive power and become insignificant.

Thus, our empirical result verifies **Hypothesis 3** that the formatting of news reports has an influence on information processing costs. After removing all inherent LF links, the LF momentum still underperforms the PE momentum significantly.

7 Further Analysis

7.1 Fundamental Predictability

A potential concern of our study is that the connected firms defined through the LF and PE link have differences in fundamental predictability. The weaker performance of LF momentum may result from its inability to capture fundamentally similar connected firms by itself; while the PE momentum is stronger because it is better at identifying connected firms with similar fundamentals. Therefore, we further compare the predictive power of the fundamentals of firms through the three links.

Specifically, we follow Ali and Hirshleifer (2020) to use the annual sales growth (*Sales growth*) and profit growth (*Profit growth*) to measure the economic fundamental conditions of a firm. $Sales\ growth_t$ is computed as the percent change of $Sales$ per share at time t to $Sales$ per share at time $t-1$; and $Profit\ growth_t$ is calculated as the difference between $Profit$ at time t and $Profit$ at time $t-1$, divided by the average of total assets at time t and $t-1$.

[22] In Table 22, we provide some examples of such LF-only links. They are mostly formed between parent firms and subsidiary firms, major shareholders and controlled firms, customers and suppliers, and acquirers and acquirees. Thus, to some extent, LF-only links could be used to represent inherent LF links.

Then we regress firms' *Sales growth* and *Profit growth* on the average growth measures of leaders (*LF sales growth*) and peers (*PE sales growth*) respectively. Consistent with Ali and Hirshleifer (2020), control variables include the firm size and book-to-market ratio. We also control for the time and entity fixed effects. Only those firms with December fiscal year ends are included. All variables are measured at the end of the calendar year and are winsorized at the 1% and 99% levels. In addition, independent variables are standardized with their cross-sectional means and standard deviations.

Table 10 presents the panel regression results. The first three columns of Panel A show that neither *LF sales growth* nor *PE sales growth* are strong predictors of future firm sales growth. Their coefficients are close, but neither is significant, indicating that the forecasting performance of the LF and PE link for fundamentals is similarly poor. Furthermore, the last three columns of Panel A show that there is a strong contemporaneous relation between firm sales and *LF sales growth* and *PE sales growth*. For instance, a one standard deviation increase in *LF sales growth* is associated with an increase of 2.482% (t-statistics = 4.09) in firm sales growth, while a one standard deviation increase in *PE sales growth* leads to an increase of 3.146% (t-statistics = 3.89) in firm sales growth. Panel B shows the same conclusions when a firm's fundamental condition is measured by profit growth instead of sales growth.

In all, the PE link is not significantly different from the LF link in either the prediction of future fundamentals or the co-movement of contemporaneous fundamentals. Therefore, there is no essential difference between the LF link and the PE link in the transmission of firms' fundamental information.

7.2 Size and the Sluggish Information Diffusion

Another concern of our study is that the weaker performance of the LF momentum may result from the slower information diffusion from small firms to big firms rather than investors' limited attention. As indicated by Hou (2007), small firms respond sluggishly to negative news from big firms, and big firms lead small firms within the same industry. Furthermore, according to statistics in Table 1, the average firm size of leader stocks is smaller than the follower stocks. So, under the framework of Hou (2007), the speed of information transmission from leader firms with a smaller size to follower firms with a larger size may have certain obstacles, which makes it more difficult for news to be transmitted from leaders to followers, thus weakening the LF momentum effect.

To verify this issue, we decompose the LF momentum into two parts: the big-leader LF momentum and the small-leader LF momentum. When constructing the predictive signal *LF_Rtn*, the big-leader LF momentum only considers leader stocks with market capitalization in the top 50% of the total sample. In contrast, the small-leader LF momentum only keeps leader stocks with market capitalization in the bottom 50% of the total sample. Then we do group portfolio sorting according to the two kinds of *LF_Rtn* and compare their long-short portfolio returns and alphas. If the weaker performance of the LF momentum is indeed due to the sluggish information diffusion from small firms to big firms, we can expect that the small-leader LF momentum underperforms the big-leader LF momentum significantly.

Table 11 reports the portfolio mean returns and alphas of the big-leader LF momentum and the small-leader LF momentum. First of all, for the equal-weighted portfolios, the small-leader LF momentum in Panel B outperforms the big-leader LF momentum in Panel A a lot. The long-short mean return and CH-4 adjusted alpha of the small-leader LF momentum are 3.88% (t-statistics = 8.41) and 3.90% (t-statistics = 8.52) per month respectively, which are almost three times of 1.27% (t-statistics = 5.31) and 1.32% (t-statistics = 5.53) of the big-leader LF momentum. Then, for the value-weighted portfolios, the two LF momentum do not show obvious discrepancies. The long-short mean return and CH-4 adjusted alpha of the small-leader LF momentum are 1.69% (t-statistics = 4.51) and 1.72% (t-statistics = 4.62) per month respectively, which are almost equal to 1.74% (t-statistics = 4.24) and 1.80% (t-statistics = 4.46) of the big-leader LF momentum.

In all, neither of the equal-weighted nor the value-weighted portfolios show a weaker performance of the small-leader LF momentum than the big-leader LF momentum. In contrast, within equal-weighted portfolios, the small-leader LF momentum is even significantly stronger than the other one. As a result, we can preclude the concern that the poorer performance of the LF momentum originates from the sluggish information diffusion from small leader firms to big follower firms.

7.3 Transaction Costs

Since the trading strategies in this paper are all at a daily frequency which is very trading intensive, it is necessary to take the impact of transaction costs on each strategy into account. Considering transaction costs can also compare the performances of the three momentum strategies in real-world investment.

We follow the method in Fan et al. (2021) to estimate the transaction cost ratio. In their study, the cost ratio in China is set to be 16 bps (buying and selling combined).[23] However, on August 28, 2023, the Ministry of Finance announced a policy to halve the stamp duty on securities transactions (from 10 bps to 5 bps). So, we re-estimate the transaction cost to be 11 bps.

Table 15 reports the portfolio performances of the LF, PE, and CM momentum after considering the 11 bps transaction cost. The PE momentum still performs the best among the three momentum strategies, with a significant positive long-short mean return and CH-4 adjusted alpha at 2.82% (t-statistics = 7.21) and 2.86% (t-statistics = 7.29) per month. However, for the LF momentum, neither the equal-weighted portfolio nor the value-weighted portfolio achieve positive returns or any risk-adjusted alphas significantly.

In summary, the PE momentum yields positive returns and alphas significantly even after considering transaction costs, while the LF momentum loses its profitability and the CM momentum underperforms the PE momentum within both equal-weighted and value-weighted portfolios. This result shows that the strong performance of PE momentum is not only theoretical but also of practical significance.

7.4 Robustness Checks

In this subsection, we conduct several further checks to test the robustness of our findings. To be specific, we first consider different identification windows for defining the LF, PE, and CM links. In addition, we also try alternative definitions of the CM link and the PE link. Finally, we adjust our trading frequency to weekly rebalancing. Our main results still hold in these robustness tests.

[23] In the Chinese A-share market, transaction costs typically have three components. Firstly, there is a stamp duty, which is levied on the total transaction amount at a rate of 10 bps (decreased to 5 bps after August 28, 2023). Note that the stamp duty is levied only on sellers. The second component is a transfer fee of 1 bps for both buying and selling transactions in the Shanghai Stock Exchange, which applies to stocks priced at 20 CNY per share. Lastly, investors pay a trading commission to brokers for executing their trades. This commission has a maximum limit of 3 bps of the transaction amount but usually ranges around 2.5 bps. Notably, institutional investors with higher trading volumes often enjoy preferential commission rates compared to individual investors. By setting the transaction cost at 16 bps, it implicitly assumes a conservative approach by considering a turnover ratio of 100% for each rebalancing period.

7.4.1 Adjustment of the Identification Window

We first re-examine our main conclusion by adjusting the identification window of each link. Specifically, in addition to the 90-day window, we also consider a shorter and a longer window of 30 and 180 days, respectively. We re-conduct the portfolio sorting analysis under the three windows.

Table 16 reports the equal-weighted portfolio sorting results of the LF, PE, and CM momentum under identification windows of 30-day, 90-day, and 180-day. Consistent with our previous conclusion, the PE momentum retains its best predictive power under all three identification windows, while the LF momentum still performs the worst. The PE momentum under the 30-day, 90-day, and 180-day generates long-short mean returns of 7.54% (t-statistics = 12.84), 5.11% (t-statistics = 12.91), and 4.10% (t-statistics = 13.01) per month, higher than 4.06% (t-statistics = 10.11), 2.66% (t-statistics = 9.62), 2.22% (t-statistics = 9.64) of the LF momentum and 6.20% (t-statistics = 14.10), 4.25% (t-statistics = 13.36), 3.58% (t-statistics = 12.96) of the CM momentum respectively. Moreover, as shown in Table 17, their differences are also statistically significant.

Interestingly, the predictive power of all three momentum decreases monotonically as the identification window becomes longer, and the differences between the three momentum effects also become smaller with the increase of the identification window. This is consistent with the findings of Ge et al. (2023) that because news articles are time-sensitive, stock relationships mined from older news often become less valid at the present. So the momentum spillover effect driven by these links is also correspondingly weakened.

Therefore, our main results are robust to the adjustment of the identification window and the PE momentum always performs better than the LF momentum as well as the CM momentum.

7.4.2 Alternative Link Specifications

In the previous analysis, we define two stocks to have a CM link if they are co-mentioned in the same sentence of a piece of news during the identification window, which is called the *same_sentence* strategy. Here we further consider the *same_article* strategy that defines two stocks to have a CM link as long as they are co-mentioned just in the same piece of news article during the identification window. Since the PE link is constructed by eliminating all LF links from CM links, the PE link also needs to be adjusted to *same_article*, while the LF link remains unchanged.

Table 18 shows the portfolio sorting performances of the LF momentum as well as the PE momentum and the CM momentum under the *same_article* strategy. The PE momentum still outperforms the LF momentum significantly after the adjustment, and the equal-weighted long-short portfolio achieves a monthly mean return and CH-4 alpha at 4.13% (t-statistics = 12.80) and 4.15% (t-statistics = 12.90) which are higher than 2.66% (t-statistics = 9.62) and 2.71% (t-statistics = 9.71) of the LF momentum. The difference between the PE and LF momentum is also statistically significant as reported in Table 19, with a long-only difference of 1.01% (t-statistics = 5.98) and a long-short difference of 1.44% (t-statistics = 6.57) per month respectively. Similarly, under the *same_article* strategy, the PE momentum continues to beat CM momentum within equal-weighted portfolios significantly, while there is no significant difference between the two in the value-weighted portfolios.

7.4.3 Weekly Rebalancing

Our previous portfolio strategy is daily trading, thus we reduce the trading frequency to weekly to see if our main results hold in non-high-frequency trading. At each trading weekend, we construct the predictive signals for the LF, PE, and CM momentum respectively which are consistent with the methods in the daily strategy. Then all sample stocks are sorted into quintiles according to their predictive signals (like LF_Rtn for the LF momentum). Stocks are equal-weighted within each quintile portfolio. The long-short strategy is buying the highest signal group and selling the lowest signal group. All portfolios are held for one week and are rebalanced weekly.

Table 20 reports the weekly performances of the LF, PE, and CM momentum under identification windows of 30 days, 90 days, and 180 days respectively. Panel A shows the portfolio mean returns, while Panel B shows the CH-4 adjusted alphas. The weekly PE momentum can still beat the LF momentum and the CM momentum under different identification windows respectively.

8 Conclusions

This paper studies how the information structure implied in news articles can influence the processing costs of investors. We separate leader-follower (LF) links and peer links (PE) from news co-mention (CM) connections. We conclude that a directed structure (LF), benefiting from featuring leader stocks in the news title, can provide a clearer linkage. This directed linkage helps investors quickly shift their attention to follower stocks in the main content, which is supported by Baidu's daily abnormal search index.

We also investigate the impact of information processing costs on market outcomes. We then conduct three momentum spillover strategies based on these three types of links. We find that, in both the portfolio sorting analysis and Fama-Macbeth regressions, the momentum spillover effect driven by the PE link significantly outperforms that of the LF and CM links. This conclusion remains robust after considering transaction costs, changing the identification windows of links, adjusting the definition of news co-mention, and reducing the frequency of trading. This finding indicates that the PE link is harder for investors to recognize than the LF or the CM link, resulting in PE momentum being stronger and more persistent than the other two types of momentum.

We finally acknowledge the role of journalists in reducing information processing costs by highlighting leader stocks in news titles. While stocks may not inherently have a leader-follower relationship, the way news is formatted can effectively depict such relationships under various circumstances, providing investors with clearer insights.

References

U. Ali and D. Hirshleifer. Shared analyst coverage: Unifying momentum spillover effects. *Journal of Financial Economics*, 136(3):649–675, June 2020. ISSN 0304405X. doi: 10.1016/j.jfineco.2019.10.007. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X19302533.

Y. Amihud. Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1):31–56, 2002. doi: 10.1016/S1386-4181(01)00024-6. URL https://www.sciencedirect.com/science/article/pii/S1386418101000246.

B. M. Barber and T. Odean. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, 21(2):785–818, Apr. 2008. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhm079. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhm079.

R. Bekkerman, E. M. Fich, and N. V. Khimich. The Effect of Innovation Similarity on Asset Prices: Evidence from Patents' Big Data. *The Review of Asset Pricing Studies*, 13(1):99–145, Feb. 2023. ISSN 2045-9920, 2045-9939. doi: 10.1093/rapstu/raac014. URL https://academic.oup.com/raps/article/13/1/99/6656364.

E. Blankespoor, E. deHaan, and C. Zhu. Capital market effects of media synthesis and dissemination: evidence from robo-journalism. *Review of Accounting Studies*, 23(1):1–36, Mar. 2018. ISSN 1380-6653, 1573-7136. doi: 10.1007/s11142-017-9422-2. URL http://link.springer.com/10.1007/s11142-017-9422-2.

E. Blankespoor, E. deHaan, and I. Marinovic. Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, 70(2-3):101344, Nov. 2020. ISSN 01654101. doi: 10.1016/j.jacceco.2020.101344. URL https://linkinghub.elsevier.com/retrieve/pii/S016541012030046X.

B. J. Bushee, J. E. Core, W. Guay, and S. J. Hamm. The Role of the Business Press as an Information Intermediary. *Journal of Accounting Research*, 48(1):1–19, Mar. 2010. ISSN 0021-8456, 1475-679X. doi: 10.1111/j.1475-679X.2009.00357.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1475-679X.2009.00357.x.

C. W. Calomiris and H. Mamaysky. How news and its context drive risk and returns around the world. *Journal of Financial Economics*, 133(2):299–336, Aug. 2019. ISSN 0304-405X. doi: 10.1016/j.jfineco.2018.11.009. URL https://www.sciencedirect.com/science/article/pii/S0304405X18303180.

J. Cao, T. Chordia, and C. Lin. Alliances and Return Predictability. *Journal of Financial and Quantitative Analysis*, 51(5):1689–1717, Oct. 2016. ISSN 0022-1090, 1756-6916. doi: 10.1017/S0022109016000600. URL https://www.cambridge.org/core/product/identifier/S0022109016000600/type/journal_article.

M. M. Carhart. On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1):57–82, Mar. 1997. ISSN 00221082. doi: 10.1111/j.1540-6261.1997.tb03808.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1997.tb03808.x.

S. Chang and D. Y. Suk. Stock Prices and the Secondary Dissemination of Information: The Wall Street Journal's "Insider Trading Spotlight" Column. *Financial Review*, 33(3):115–128, Aug. 1998. ISSN 0732-

8516, 1540-6288. doi: 10.1111/j.1540-6288.1998.tb01386.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6288.1998.tb01386.x.

J. Chen, G. Tang, J. Yao, and G. Zhou. Investor Attention and Stock Returns. *Journal of Financial and Quantitative Analysis*, 57(2):455–484, Mar. 2022. ISSN 0022-1090, 1756-6916. doi: 10.1017/S0022109021000090. URL https://www.cambridge.org/core/product/identifier/S0022109021000090/type/journal_article.

X. Chen, W. He, L. Tao, and J. Yu. Attention and Underreaction-Related Anomalies. *Management Science*, 69(1):636–659, Jan. 2023. ISSN 0025-1909, 1526-5501. doi: 10.1287/mnsc.2022.4332. URL https://pubsonline.informs.org/doi/10.1287/mnsc.2022.4332.

L. Cohen and A. Frazzini. Economic Links and Predictable Returns. *The Journal of Finance*, 63(4):1977–2011, Aug. 2008. ISSN 00221082. doi: 10.1111/j.1540-6261.2008.01379.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2008.01379.x.

Z. Da, J. Engelberg, and P. Gao. In Search of Attention. *The Journal of Finance*, 66(5):1461–1499, Oct. 2011. ISSN 00221082. doi: 10.1111/j.1540-6261.2011.01679.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2011.01679.x.

S. Dellavigna and J. M. Pollet. Investor Inattention and Friday Earnings Announcements. *The Journal of Finance*, 64(2):709–749, Apr. 2009. ISSN 0022-1082, 1540-6261. doi: 10.1111/j.1540-6261.2009.01447.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2009.01447.x.

C. Dougal, J. Engelberg, D. García, and C. A. Parsons. Journalists and the Stock Market. *Review of Financial Studies*, 25(3):639–679, Mar. 2012. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhr133. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhr133.

Q. Du, D. Liang, Z. Chen, and J. Tu. Concept links and return momentum. *Journal of Banking & Finance*, 134:106329, Jan. 2022. ISSN 03784266. doi: 10.1016/j.jbankfin.2021.106329. URL https://linkinghub.elsevier.com/retrieve/pii/S0378426621002806.

B. Duan, R. Wnag, and R. Zhang. Economic Links and Stock Returns in Chinese A-Share Market. *Journal of Financial Research*, (2):171–188, 2022. ISSN 1002-7246. URL https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFDAUTO&filename=JRYJ202202010&v=.

A. Eisdorfer, K. Froot, G. Ozik, and R. Sadka. Competition Links and Stock Returns. *The Review of Financial Studies*, 35(9):4300–4340, Aug. 2022. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhab133. URL https://academic.oup.com/rfs/article/35/9/4300/6470574.

J. E. Engelberg and C. A. Parsons. The Causal Impact of Media in Financial Markets. *The Journal of Finance*, 66(1):67–97, Feb. 2011. ISSN 0022-1082, 1540-6261. doi: 10.1111/j.1540-6261.2010.01626.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01626.x.

E. F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2):383, May 1970. ISSN 00221082. doi: 10.2307/2325486. URL https://www.jstor.org/stable/2325486?origin=crossref.

E. F. Fama and K. R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, Feb. 1993. ISSN 0304405X. doi: 10.1016/0304-405X(93)90023-5. URL https://linkinghub.elsevier.com/retrieve/pii/0304405X93900235.

E. F. Fama and K. R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, Apr. 2015. ISSN 0304405X. doi: 10.1016/j.jfineco.2014.10.010. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X14002323.

E. F. Fama and J. D. MacBeth. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3):607–636, May 1973. ISSN 0022-3808, 1537-534X. doi: 10.1086/260061. URL https://www.journals.uchicago.edu/doi/10.1086/260061.

J. Fan, L. Xue, and Y. Zhou. How Much Can Machines Learn Finance From Chinese Text Data? *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3765862. URL https://www.ssrn.com/abstract=3765862.

S. Ge, S. Li, and O. Linton. News-implied linkages and local dependency in the equity market. *Journal of Econometrics*, page S0304407622001488, Sept. 2022. ISSN 03044076. doi: 10.1016/j.jeconom.2022.07.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0304407622001488.

S. Ge, S. Li, and H. Zheng. Diamond Cuts Diamond: News Co-mention Momentum Spillover Prevails in China. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4489005. URL https://www.ssrn.com/abstract=4489005.

S. Gervais, R. Kaniel, and D. H. Mingelgrin. The High-Volume Return Premium. *The Journal of Finance*, 56(3):877–919, June 2001. ISSN 0022-1082, 1540-6261. doi: 10.1111/0022-1082.00349. URL https://onlinelibrary.wiley.com/doi/10.1111/0022-1082.00349.

L. Guo, L. Peng, Y. Tao, and J. Tu. Joint News, Attention Spillover,and Market Returns. *SSRN Electronic Journal*, 2017. ISSN 1556-5068. doi: 10.2139/ssrn.2927561. URL https://ssrn.com/abstract=2927561.

A. Hillert, H. Jacobs, and S. Müller. Media Makes Momentum. *Review of Financial Studies*, 27(12):3467–3501, Dec. 2014. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhu061. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/hhu061.

D. Hirshleifer and S. H. Teoh. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1-3):337–386, Dec. 2003. ISSN 01654101. doi: 10.1016/j.jacceco.2003.10.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0165410103000648.

D. Hirshleifer, S. S. Lim, and S. H. Teoh. Limited Investor Attention and Stock Market Misreactions to Accounting Information. *Review of Asset Pricing Studies*, 1(1):35–73, Dec. 2011. ISSN 2045-9920, 2045-9939. doi: 10.1093/rapstu/rar002. URL https://academic.oup.com/raps/article-lookup/doi/10.1093/rapstu/rar002.

D. Hirshleifer, P.-H. Hsu, and D. Li. Innovative efficiency and stock returns. *Journal of Financial Economics*, 107(3):632–654, Mar. 2013. ISSN 0304405X. doi: 10.1016/j.jfineco.2012.09.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X12001961.

G. Hoberg and G. Phillips. Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy*, 124(5):1423–1465, Oct. 2016. ISSN 0022-3808, 1537-534X. doi: 10.1086/688176. URL https://www.journals.uchicago.edu/doi/10.1086/688176.

G. Hoberg and G. M. Phillips. Text-Based Industry Momentum. *Journal of Financial and Quantitative Analysis*, 53(6):2355–2388, Dec. 2018. ISSN 0022-1090, 1756-6916. doi: 10.1017/S0022109018000479. URL https://www.cambridge.org/core/product/identifier/S0022109018000479/type/journal_article.

K. Hou. Industry Information Diffusion and the Lead-lag Effect in Stock Returns. *Review of Financial Studies*, 20(4):1113–1138, July 2007. ISSN 0893-9454, 1465-7368. doi: 10.1093/revfin/hhm003. URL https://academic.oup.com/rfs/article-lookup/doi/10.1093/revfin/hhm003.

K. Hou, W. Xiong, and L. Peng. A Tale of Two Anomalies: The Implications of Investor Attention for Price and Earnings Momentum. *SSRN Electronic Journal*, 2009. ISSN 1556-5068. doi: 10.2139/ssrn.976394. URL http://www.ssrn.com/abstract=976394.

K. Hou, F. Qiao, and X. Zhang. Finding Anomalies in China. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4322815. URL https://www.ssrn.com/abstract=4322815.

J. Hu and W. K. Härdle. Networks of News and the Cross-Sectional Returns. *SSRN Electronic Journal*, 2021. ISSN 1556-5068. doi: 10.2139/ssrn.3904012. URL https://www.ssrn.com/abstract=3904012.

S. Huang, T.-C. Lin, and H. Xiang. Psychological barrier and cross-firm return predictability. *Journal of Financial Economics*, 142(1):338–356, Oct. 2021. ISSN 0304405X. doi: 10.1016/j.jfineco.2021.06.006. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X21002725.

S. Huang, C. M. Lee, Y. Song, and H. Xiang. A frog in every pan: Information discreteness and the lead-lag returns puzzle. *Journal of Financial Economics*, 145(2):83–102, Aug. 2022. ISSN 0304405X. doi: 10.1016/j.jfineco.2021.10.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X21004761.

G. Huberman and T. Regev. Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar. *The Journal of Finance*, 56(1):387–396, Feb. 2001. ISSN 0022-1082, 1540-6261. doi: 10.1111/0022-1082.00330. URL https://onlinelibrary.wiley.com/doi/10.1111/0022-1082.00330.

Z. Jin and F. W. Li. Geographic links and predictable returns. *Working Paper*, 2020. URL https://ink.library.smu.edu.sg/lkcsb_research/6593.

B. Julesz. A brief outline of the texton theory of human vision. *Trends in Neurosciences*, 7(2):41–45, Feb. 1984. ISSN 01662236. doi: 10.1016/S0166-2236(84)80275-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0166223684802751.

A. Lawrence, J. Ryans, E. Sun, and N. Laptev. Earnings announcement promotions: A Yahoo Finance field experiment. *Journal of Accounting and Economics*, 66(2-3):399–414, Nov. 2018. ISSN 01654101. doi: 10.1016/j.jacceco.2018.08.004. URL https://linkinghub.elsevier.com/retrieve/pii/S0165410118300442.

C. M. Lee, S. T. Sun, R. Wang, and R. Zhang. Technological links and predictable returns. *Journal of Financial Economics*, 132(3):76–96, June 2019. ISSN 0304405X. doi: 10.1016/j.jfineco.2018.11.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X18303167.

J. Li and J. Yu. Investor attention, psychological anchors, and stock return predictability. *Journal of Financial Economics*, 104(2):401–419, May 2012. ISSN 0304405X. doi: 10.1016/j.jfineco.2011.04.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X11002121.

J. Liu, R. F. Stambaugh, and Y. Yuan. Size and value in China. *Journal of Financial Economics*, 134(1):48–69, Oct. 2019. ISSN 0304405X. doi: 10.1016/j.jfineco.2019.03.008. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X19300625.

L. Menzly and O. Ozbas. Market Segmentation and Cross-predictability of Returns. *The Journal of Finance*, 65(4):1555–1580, Aug. 2010. ISSN 00221082. doi: 10.1111/j.1540-6261.2010.01578.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2010.01578.x.

T. J. Moskowitz and M. Grinblatt. Do Industries Explain Momentum? *The Journal of Finance*, 54(4):1249–1290, Aug. 1999. ISSN 00221082. doi: 10.1111/0022-1082.00146. URL http://doi.wiley.com/10.1111/0022-1082.00146.

W. K. Newey and K. D. West. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3):703, May 1987. ISSN 00129682. doi: 10.2307/1913610. URL https://www.jstor.org/stable/1913610?origin=crossref.

K. Obaid and K. Pukthuanthong. A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, 144(1):273–297, Apr. 2022. ISSN 0304-405X. doi: 10.1016/j.jfineco.2021.06.002. URL https://www.sciencedirect.com/science/article/pii/S0304405X21002683.

C. A. Parsons, R. Sabbatucci, and S. Titman. Geographic Lead-Lag Effects. *The Review of Financial Studies*, 33 (10):4721–4770, Oct. 2020. ISSN 0893-9454, 1465-7368. doi: 10.1093/rfs/hhz145. URL https://academic.oup.com/rfs/article/33/10/4721/5682420.

L. Peng and W. Xiong. Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3):563–602, June 2006. ISSN 0304405X. doi: 10.1016/j.jfineco.2005.05.003. URL https://linkinghub.elsevier.com/retrieve/pii/S0304405X05002138.

A. Scherbina and B. Schlusche. Economic Linkages Inferred from News Stories and the Predictability of Stock Returns. *SSRN Electronic Journal*, 2013. ISSN 1556-5068. doi: 10.2139/ssrn.2363436. URL http://www.ssrn.com/abstract=2363436.

G. Schwenkler and H. Zheng. The Network of Firms Implied by the News. *SSRN Electronic Journal*, 2019. ISSN 1556-5068. doi: 10.2139/ssrn.3320859. URL https://www.ssrn.com/abstract=3320859.

G. Schwenkler and H. Zheng. News-Driven Peer Co-Movement in Crypto Markets. *Available at SSRN 3572471*, page 61, Sept. 2021. doi: 10.2139/ssrn.3572471. URL https://ssrn.com/abstract=3572471.

A. Tversky and D. Kahneman. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131, Sept. 1974. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.185.4157.1124. URL https://www.science.org/doi/10.1126/science.185.4157.1124.

H. Wang. News Link and Predictable Returns. *SSRN Electronic Journal*, 2023. ISSN 1556-5068. doi: 10.2139/ssrn.4458612. URL https://www.ssrn.com/abstract=4458612.

Tables

Table 1: Descriptive statistics

Panel A: leader stocks								
	count	mean	std	min	25%	50%	75%	max
Float_Size (billion)	862831	7.79	12.47	0.15	2.39	4.31	8.30	408.54
Total_Size (billion)	862831	10.11	15.37	0.35	3.43	5.78	11.04	413.97
Turnover	862831	0.03	0.04	0	0.01	0.02	0.03	0.95
Rtn (%)	862831	0.11	3.95	-26.33	-1.44	0	1.46	1975.36
Liquidity	862596	0.64	93.43	0	0.01	0.02	0.05	30043.77
Beta	842385	1.09	0.40	-19.22	0.84	1.11	1.34	3.78
ROE	852058	0.00	1.71	-207.40	0.01	0.03	0.07	5.32
BM	859743	0.44	0.35	0.00	0.21	0.36	0.57	9.95
ASVI_Code	740020	0.30	2.42	-1.00	-0.07	0.12	0.41	1492.06
ASVI_All	823854	0.31	3.35	-1.00	-0.08	0.11	0.41	2401.43
Panel B: follower stocks								
	count	mean	std	min	25%	50%	75%	max
Float_Size (billion)	318119	12.00	45.35	0.12	2.49	4.82	9.89	1337.48
Total_Size (billion)	318119	14.55	46.50	0.19	3.40	6.31	12.89	1337.48
Turnover	318119	0.02	0.03	0.00	0.01	0.01	0.03	0.90
Rtn (%)	318119	0.07	4.47	-36.90	-1.35	0	1.37	1677.26
Liquidity	318090	0.17	6.67	0	0.01	0.02	0.06	2186.77
Beta	313393	1.10	0.40	-2.74	0.86	1.12	1.34	3.49
ROE	313612	0.00	2.13	-176.38	0.01	0.03	0.07	204.69
BM	314980	0.48	0.38	0.00	0.21	0.38	0.62	5.44
ASVI_Code	294312	0.24	1.14	-1.00	-0.10	0.09	0.38	291.13
ASVI_All	312099	0.23	0.75	-1.00	-0.11	0.07	0.35	74.56
Panel C: peer stocks								
	count	mean	std	min	25%	50%	75%	max
Float_Size (billion)	319534	8.33	14.34	0.09	2.26	4.22	8.51	401.72
Total_Size (billion)	319534	10.50	16.30	0.13	3.23	5.56	11.22	401.72
Turnover	319534	0.03	0.04	0	0.01	0.01	0.03	0.81
Rtn (%)	319534	0.10	3.19	-20.18	-1.41	0	1.42	286.33
Liquidity	319430	2.55	233.38	0	0.01	0.03	0.06	57977.27
Beta	311513	1.07	0.42	-0.73	0.81	1.09	1.34	3.54
ROE	314546	0.02	0.72	-53.04	0.01	0.03	0.07	7.81
BM	317601	0.46	0.37	0.00	0.21	0.37	0.60	9.79
ASVI_Code	275032	0.26	0.80	-1.00	-0.07	0.12	0.39	94.83
ASVI_All	303224	0.27	1.04	-1.00	-0.08	0.10	0.38	264.79
Panel D: all sample stocks								
	count	mean	std	min	25%	50%	75%	max
Float_Size	7251234	12.46	55.16	0.06	2.07	4.00	8.48	3267.37
Total_Size	7251234	15.37	57.36	0.07	3.18	5.54	11.39	3267.37
Turnover	7251234	0.03	0.04	0	0.01	0.01	0.03	0.95
Rtn	7251234	0.10	3.85	-47.39	-1.40	0	1.42	1975.36
Liquidity	7249251	0.48	69.45	0	0.01	0.03	0.06	57977.27
Beta	7017974	1.12	0.40	-19.22	0.87	1.13	1.36	10.06
ROE	7145286	0.03	1.98	-207.40	0.01	0.03	0.07	281.99
BM	7213212	0.44	0.36	0.00	0.21	0.35	0.57	22.04
ASVI_Code	6177076	0.28	1.94	-1.00	-0.08	0.12	0.41	1492.06
ASVI_All	6847125	0.29	2.23	-1.00	-0.09	0.10	0.39	2401.43

This table reports summary statistics. The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. Panels A to C report the summary statistics for leader, follower, and peer stocks respectively; and Panel D reports the full sample statistics. At each trading day, if one stock has only been identified as the leader stock during the previous 90 days, then it is called the leader stock. The definitions of follower and peer stocks are similar. *ASVI_Code* is the Baidu abnormal search volume index (Guo et al., 2017) with only stock codes as search keywords; while *ASVI_All* is the index with codes, Chinese abbreviation names, and full names as keywords. For the definitions of other variables see subsection 3.2.

Table 2: Characteristic differences between leaders, followers, and peers

	Leaders minus Followers	Leaders minus Peers	Followers minus Peers
Float_Size	-4.2148	-0.5400	3.6748
	(-51.70)	(-18.81)	(43.58)
Total_Size	-4.4443	-0.3886	4.0557
	(-52.85)	(-11.69)	(46.43)
Turnover	0.0048	0.0018	-0.0030
	(62.54)	(20.66)	(-31.76)
Rtn (%)	0.0371	0.0030	-0.0341
	(4.13)	(0.43)	(-3.50)
Liquidity	0.4737	-1.9092	-2.3829
	(4.68)	(-4.49)	(-5.77)
Beta	-0.0125	0.0175	0.0300
	(-14.99)	(19.99)	(28.87)
ROE	-0.0044	-0.0217	-0.0173
	(-1.03)	(-9.62)	(-4.32)
BM	-0.0341	-0.0154	0.0188
	(-43.75)	(-20.44)	(19.93)
ASVI_Code	0.0580	0.0405	-0.0175
	(16.50)	(12.68)	(-6.72)
ASVI_All	0.0815	0.0397	-0.0418
	(20.79)	(9.59)	(-18.12)

The table shows the t-test results of the characteristic differences among leader, follower, and peer stocks. The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. At each trading day, if one stock has only been identified as the leader stock during the previous 90 days, then it is called the leader stock. The definitions of follower and peer stocks are similar. $ASVI\_Code$ is the Baidu abnormal search volume index (Guo et al., 2017) with only stock codes as search keywords; while $ASVI\_All$ is the index with codes, firm names, and short names as keywords. For the definitions of other variables see subsection 3.2.

Table 3: Panel regressions of investors' attention and LF, PE events

Panel A: Single pair of event dummies						
n	LF_Event	PE_Event	Control	Avg. R2	Total Obs.	Avg. Obs.
-10	0.0132	0.0161	YES	0.046	6584054	2708
	(1.90)	(2.15)				
-9	0.0141	0.0189	YES	0.046	6584054	2708
	(1.98)	(2.48)				
-8	0.0153	0.0201	YES	0.046	6584054	2708
	(2.09)	(2.70)				
-7	0.0175	0.0193	YES	0.046	6584054	2708
	(2.30)	(2.59)				
-6	0.0214	0.0202	YES	0.046	6584054	2708
	(2.69)	(2.75)				
-5	0.0256	0.0229	YES	0.046	6584054	2708
	(3.16)	(3.31)				
-4	0.0286	0.0240	YES	0.046	6584054	2708
	(3.45)	(3.69)				
-3	0.0316	0.0266	YES	0.046	6584054	2708
	(3.73)	(3.54)				
-2	0.0342	0.0285	YES	0.046	6584054	2708
	(4.05)	(3.88)				
-1	0.0371	0.0292	YES	0.046	6584054	2708
	(4.40)	(4.22)				
0	**0.0394**	**0.0283**	**YES**	**0.046**	**6584054**	**2708**
	(4.59)	**(4.39)**				
1	0.0387	0.0269	YES	0.046	6584054	2708
	(4.48)	(4.39)				
2	0.0386	0.0284	YES	0.046	6584054	2708
	(4.50)	(4.61)				
3	0.0388	0.0292	YES	0.046	6584054	2708
	(4.54)	(4.57)				
4	0.0383	0.0266	YES	0.046	6584054	2708
	(4.51)	(4.33)				
5	0.0381	0.0269	YES	0.046	6584054	2708
	(4.38)	(4.31)				
6	0.0373	0.0279	YES	0.046	6584054	2708
	(4.28)	(4.44)				
7	0.0371	0.0288	YES	0.046	6584054	2708
	(4.26)	(4.49)				
8	0.0374	0.0262	YES	0.046	6584054	2708
	(4.28)	(4.19)				
9	0.0374	0.0259	YES	0.046	6584054	2708
	(4.26)	(4.17)				
10	0.0357	0.0266	YES	0.046	6584054	2708
	(4.10)	(4.29)				

Panel B: All pairs of event dummies		
n	LF_Event	PE_Event
-10	-0.0045	-0.0054
	(-1.63)	(-1.41)
-9	-0.0038	0.0027
	(-2.05)	(0.47)
-8	-0.0047	0.0037
	(-2.26)	(1.17)
-7	-0.0044	-0.0028
	(-1.74)	(-0.57)
-6	-0.0006	-0.0032
	(-0.27)	(-0.87)
-5	0.0028	0.0026
	(1.45)	(1.08)
-4	0.0030	-0.0004
	(1.42)	(-0.08)
-3	0.0040	0.0044
	(1.62)	(0.96)
-2	0.0042	0.0063
	(2.10)	(1.93)
-1	0.0068	0.0071
	(3.58)	(2.51)
0	**0.0102**	**0.0036**
	(5.15)	**(1.48)**
1	0.0054	-0.0021
	(2.42)	(-0.34)
2	0.0047	0.0039
	(2.67)	(1.41)
3	0.0051	0.0078
	(3.08)	(1.83)
4	0.0045	-0.0024
	(2.40)	(-0.54)
5	0.0048	0.0004
	(2.64)	(0.16)
6	0.0030	0.0034
	(1.87)	(1.88)
7	0.0033	0.0086
	(2.11)	(1.80)
8	0.0057	-0.0007
	(3.47)	(-0.17)
9	0.0079	0.0009
	(3.79)	(0.48)
10	0.0079	0.0079
	(2.37)	(3.06)
Control	YES	
Avg. R2	0.046	
Total Obs.	6584054	
Avg. Obs.	2708	

This table reports the panel results of regressing Baidu abnormal search volume $ASVI$ of stocks on dummy variables indicating whether leaders or peers have attention-grabbing events. The dependent variable of all regressions is the daily abnormal search index ($ASVI$) which is computed as the percentage change between the daily search volume index (SVI) for a stock and its past 1-year mean (Guo et al., 2017), skipping the most recent day. The two two dummy variables are $LF\_Events_{i,t}$ and $PE\_Events_{i,t}$, taking 1 if the $ASVI$ of any of the leaders/peers of stock i ranks in the top 10% of all samples on trading day t, otherwise, taking 0. In Panel A, there are in total 21 regression results. The core explanatory variables are $LF\_Events_{i,t+n}$ and $PE\_Events_{i,t+n}$ where n is an integer from -10 to 10. Panel B shows one regression result where all attention-grabbing event dummy variables (i.e., n from -10 to 10) are added. Control variables include the stock daily return, stock size (taking logarithms), and the daily turnover rate. All regression control for the time and entity effects. T-statistics based on standard errors clustered by entity and time are shown in parentheses. The coefficients of the dummy variables with $n = 0$ are marked in bold.

Table 4: Portfolio sorting performances of LF, PE, and CM momentum

Panel A: LF momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	1.01	0.93	0.87	0.95	0.58	0.19	0.22	0.20	0.20	-0.09
	(1.25)	(1.21)	(1.14)	(1.23)	(0.75)	(0.29)	(0.32)	(0.29)	(0.29)	(-0.13)
2	1.37	1.28	1.24	1.32	0.94	0.71	0.72	0.69	0.73	0.40
	(1.75)	(1.73)	(1.68)	(1.75)	(1.25)	(1.11)	(1.14)	(1.10)	(1.14)	(0.63)
3	1.56	1.48	1.44	1.52	1.15	1.08	1.08	1.04	1.11	0.75
	(2.02)	(2.03)	(1.98)	(2.05)	(1.56)	(1.76)	(1.82)	(1.75)	(1.84)	(1.23)
4	1.87	1.78	1.75	1.83	1.48	1.62	1.64	1.62	1.66	1.40
	(2.41)	(2.46)	(2.42)	(2.47)	(2.01)	(2.58)	(2.72)	(2.65)	(2.72)	(2.24)
5	3.69	3.60	3.56	3.64	3.30	2.07	2.07	2.05	2.10	1.83
	(4.50)	(4.70)	(4.68)	(4.66)	(4.27)	(2.95)	(3.09)	(3.07)	(3.08)	(2.69)
5-1	**2.66**	**2.65**	**2.66**	**2.66**	**2.71**	**1.88**	**1.85**	**1.85**	**1.90**	**1.92**
	(9.62)	**(9.42)**	**(9.50)**	**(9.46)**	**(9.71)**	**(4.49)**	**(4.35)**	**(4.36)**	**(4.42)**	**(4.72)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: PE momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.23	0.15	0.10	0.17	-0.18	-0.43	-0.41	-0.44	-0.43	-0.72
	(0.28)	(0.19)	(0.13)	(0.21)	(-0.23)	(-0.62)	(-0.60)	(-0.65)	(-0.62)	(-1.05)
2	1.09	1.00	0.95	1.03	0.66	0.70	0.71	0.65	0.72	0.35
	(1.36)	(1.32)	(1.27)	(1.34)	(0.87)	(1.05)	(1.09)	(1.00)	(1.09)	(0.54)
3	1.80	1.72	1.67	1.76	1.38	1.14	1.14	1.10	1.16	0.84
	(2.29)	(2.32)	(2.26)	(2.32)	(1.84)	(1.78)	(1.84)	(1.77)	(1.84)	(1.31)
4	2.15	2.07	2.03	2.12	1.75	1.90	1.93	1.89	1.96	1.65
	(2.71)	(2.79)	(2.75)	(2.79)	(2.31)	(2.96)	(3.10)	(3.01)	(3.11)	(2.58)
5	5.34	5.24	5.20	5.29	4.96	3.02	3.03	3.01	3.06	2.80
	(6.33)	(6.68)	(6.65)	(6.61)	(6.23)	(4.28)	(4.47)	(4.43)	(4.45)	(4.03)
5-1	**5.11**	**5.09**	**5.09**	**5.12**	**5.14**	**3.46**	**3.46**	**3.47**	**3.50**	**3.54**
	(12.91)	**(12.58)**	**(12.77)**	**(12.64)**	**(12.97)**	**(7.10)**	**(6.81)**	**(6.91)**	**(6.91)**	**(7.25)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: CM momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.61	0.52	0.47	0.54	0.20	-0.36	-0.35	-0.38	-0.36	-0.66
	(0.74)	(0.66)	(0.60)	(0.69)	(0.25)	(-0.51)	(-0.50)	(-0.55)	(-0.51)	(-0.94)
2	1.13	1.03	0.99	1.07	0.68	0.34	0.33	0.29	0.34	-0.01
	(1.41)	(1.37)	(1.32)	(1.39)	(0.89)	(0.51)	(0.50)	(0.45)	(0.51)	(-0.01)
3	1.70	1.60	1.56	1.65	1.27	1.40	1.41	1.35	1.44	1.05
	(2.16)	(2.17)	(2.12)	(2.18)	(1.70)	(2.16)	(2.23)	(2.13)	(2.24)	(1.63)
4	2.07	1.98	1.94	2.03	1.66	1.91	1.92	1.89	1.95	1.64
	(2.60)	(2.67)	(2.61)	(2.67)	(2.20)	(2.87)	(3.01)	(2.92)	(3.00)	(2.48)
5	4.88	4.77	4.72	4.82	4.47	3.04	3.02	3.01	3.05	2.78
	(5.86)	(6.17)	(6.13)	(6.10)	(5.71)	(4.32)	(4.55)	(4.53)	(4.52)	(4.11)
5-1	**4.25**	**4.23**	**4.24**	**4.25**	**4.27**	**3.41**	**3.39**	**3.41**	**3.42**	**3.47**
	(13.36)	**(13.06)**	**(13.24)**	**(13.09)**	**(13.39)**	**(7.43)**	**(7.13)**	**(7.16)**	**(7.21)**	**(7.49)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of the cross-firm momentum driven by LF, PE, and CM links, respectively. The LF link defines a directional relationship from the stock appearing in the news title (leader) to the stock mentioned in the body text of the same news (follower) during the previous 90 days. The CM link considers stocks appearing in the same sentence of a news article during the previous 90 days as news comention connected stocks. The PE link defines two firms to be peer stocks with each other if they are only mentioned in the same news body during the previous 90 days. The predictive signal of one focal stock is the average return of its connected stocks defined by each of the three links, weighted by the co-mention times (for LF momentum, the weight is the leading times). For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. We report the mean return of each portfolio, as well as returns adjusted by the Fama and French (1993) three-factor model, Fama and French (2015) five-factor model, Carhart (1997) four-factor model, and Liu et al. (2019) Chinese four-factor model. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 5: Portfolio return spreads between LF, PE, and CM momentum

Panel A: Equal-weighted		
	Long-only difference	Long-short difference
PE minus LF	**1.60**	**2.39**
	(8.03)	**(9.05)**
PE minus CM	**0.44**	**0.83**
	(3.99)	**(5.94)**
LF minus CM	**-1.14**	**-1.53**
	(-7.61)	**(-7.95)**
Panel B: Value-weighted		
	Long-only difference	Long-short difference
PE minus LF	**0.93**	**1.56**
	(3.68)	**(4.55)**
PE minus CM	-0.02	0.04
	(-0.12)	(0.19)
LF minus CM	**-0.94**	**-1.49**
	(-4.65)	**(-5.25)**

The table shows statistical test results of long-only and long-short portfolio return spreads between the PE momentum, LF momentum, and CM momentum. The return spread is computed by taking the difference between the time series returns of two portfolios. All daily returns and alphas are converted into monthly percentages using compound interest. The sample period is 2012 - 2021. Newey and West (1987) adjusted t-statistics are shown in parentheses. Return spreads with t-statistics higher than 2.00 are highlighted in bold.

Table 6: Fama-MacBeth Regressions

	1	2	3	4	5	6	7
LF_Rtn	0.0076			0.0040		0.0013	0.0029
	(2.32)			(1.09)		(0.41)	(0.74)
PE_Rtn		0.0313		0.0214	0.0223		0.0180
		(6.93)		(4.02)	(3.84)		(2.73)
CM_Rtn			0.0240		0.0101	0.0174	0.0035
			(6.83)		(1.78)	(3.86)	(0.52)
Size	-0.0809	-0.0858	-0.0922	-0.0765	-0.0854	-0.0807	-0.0763
	(-10.35)	(-10.73)	(-11.72)	(-9.49)	(-10.68)	(-10.16)	(-9.46)
BM	0.0086	0.0120	0.0094	0.0135	0.0120	0.0107	0.0134
	(1.47)	(1.94)	(1.59)	(2.13)	(1.94)	(1.75)	(2.12)
EP	0.0014	0.0024	0.0037	-0.0010	0.0023	0.0006	-0.0009
	(0.34)	(0.52)	(0.85)	(-0.20)	(0.49)	(0.13)	(-0.19)
ROE	0.0429	0.0467	0.0456	0.0473	0.0467	0.0461	0.0471
	(8.31)	(8.77)	(9.10)	(8.43)	(8.77)	(8.54)	(8.40)
Rtn	0.2012	0.1968	0.2097	0.1882	0.1955	0.1999	0.1872
	(18.59)	(18.33)	(19.90)	(17.07)	(18.27)	(18.34)	(17.01)
Assets_growth	0.0372	0.0363	0.0371	0.0334	0.0359	0.0373	0.0331
	(9.64)	(9.83)	(10.63)	(8.09)	(9.78)	(9.57)	(8.03)
Turnover	-0.1498	-0.1475	-0.1567	-0.1370	-0.1475	-0.1473	-0.1368
	(-15.41)	(-15.49)	(-17.10)	(-13.50)	(-15.49)	(-15.08)	(-13.48)
Intercept	0.0998	0.1027	0.1013	0.1016	0.1027	0.1005	0.1017
	(2.43)	(2.50)	(2.48)	(2.46)	(2.49)	(2.45)	(2.47)
Avg. R Square	0.0943	0.0968	0.0885	0.1078	0.0976	0.0992	0.1088
Total Obs.	1955985	1965781	2597936	1467118	1965781	1728256	1467118
Avg. Obs.	805	809	1069	604	809	711	604

This table reports the results of Fama and MacBeth (1973) regressions. The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. The dependent variable is the focal stock return on the next trading day; while independent variables are *LF_Rtn*, *PE_Rtn*, and *CM_Rtn* which are the connected stock returns by the LF, PE, and CM link of the focal stock respectively. The control variables include firm size (*Size*, taking logarithms), book-to-market ratio (*BM*), earnings-to-price ratio (*EP*), return on equity (*ROE*), past daily return (*Rtn*), total assets growth (*Assets_growth*), and daily turnover rate (*Turnover*). Non-return variables are winsorized at 1% and 99% in the cross-section and all independent variables are cross-sectionally standardized. Newey and West (1987) adjusted t-statistics are shown in parentheses. For brevity, all coefficients are shown multiplied by 100.

Table 7: Predictability over longer horizons

Dependent variable		*Rtn* (t+2)			*Rtn* (t+3)			*Rtn* (t+4)			*Rtn* (t+5)	
	1	2	3	4	5	6	7	8	9	10	11	12
LF_Rtn	0.0170			0.0165			0.0097			0.0100		
	(5.33)			(5.31)			(3.25)			(3.30)		
PE_Rtn		0.0358			0.0301			0.0298			0.0255	
		(7.69)			(6.61)			(6.61)			(5.79)	
CM_Rtn			0.0323			0.0281			0.0228			0.0212
			(8.98)			(8.46)			(6.60)			(6.14)
Size	-0.0735	-0.0771	-0.0838	-0.0655	-0.0691	-0.0748	-0.0644	-0.0696	-0.0738	-0.0582	-0.0617	-0.0661
	(-8.96)	(-9.28)	(-10.27)	(-8.26)	(-8.53)	(-9.46)	(-8.42)	(-8.88)	(-9.66)	(-7.61)	(-7.99)	(-8.76)
BM	0.0135	0.0180	0.0156	0.0148	0.0180	0.0173	0.0141	0.0174	0.0171	0.0146	0.0181	0.0175
	(2.15)	(2.72)	(2.44)	(2.39)	(2.73)	(2.74)	(2.23)	(2.59)	(2.66)	(2.29)	(2.69)	(2.71)
EP	-0.0006	-0.0006	0.0014	-0.0017	-0.0008	0.0004	-0.0033	-0.0022	-0.0019	-0.0049	-0.0044	-0.0037
	(-0.14)	(-0.13)	(0.30)	(-0.36)	(-0.15)	(0.08)	(-0.71)	(-0.44)	(-0.41)	(-1.07)	(-0.86)	(-0.79)
ROE	0.0459	0.0490	0.0487	0.0461	0.0497	0.0500	0.0452	0.0488	0.0496	0.0453	0.0493	0.0499
	(8.68)	(8.92)	(9.39)	(8.63)	(9.02)	(9.46)	(8.43)	(8.89)	(9.43)	(8.32)	(8.98)	(9.45)
Rtn	0.0434	0.0425	0.0530	0.0438	0.0464	0.0500	0.0368	0.0339	0.0398	-0.0327	-0.0328	-0.0269
	(4.95)	(5.07)	(6.32)	(5.86)	(6.46)	(7.01)	(5.10)	(4.68)	(5.64)	(-4.86)	(-5.01)	(-4.27)
Assets_growth	0.0336	0.0330	0.0337	0.0306	0.0289	0.0298	0.0301	0.0280	0.0287	0.0277	0.0261	0.0272
	(8.46)	(8.81)	(9.54)	(7.79)	(7.87)	(8.69)	(7.71)	(7.68)	(8.49)	(7.32)	(7.37)	(8.28)
Turnover	-0.1066	-0.1062	-0.1144	-0.0844	-0.0846	-0.0902	-0.0839	-0.0858	-0.0889	-0.0720	-0.0742	-0.0765
	(-11.43)	(-11.56)	(-12.96)	(-9.75)	(-9.97)	(-11.22)	(-10.04)	(-10.57)	(-11.47)	(-8.78)	(-9.36)	(-10.14)
Intercept	0.0961	0.0976	0.0965	0.0930	0.0940	0.0923	0.0895	0.0909	0.0884	0.0841	0.0860	0.0840
	(2.36)	(2.39)	(2.37)	(2.30)	(2.32)	(2.29)	(2.23)	(2.26)	(2.21)	(2.10)	(2.14)	(2.11)
Avg. R Square	0.0858	0.0878	0.0794	0.0821	0.0838	0.0757	0.0795	0.0810	0.0731	0.0777	0.0782	0.0706
Total Obs.	1955984	1965780	2597934	1955983	1965779	2597932	1955982	1965778	2597930	1955981	1965777	2597928
Avg. Obs.	805	809	1069	805	809	1069	805	809	1069	805	809	1069

This table reports the results of Fama and MacBeth (1973) regressions over longer horizons. The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. The independent variables are *LF_Rtn*, *PE_Rtn*, and *CM_Rtn* which are the connected stock returns by the LF, PE, and CM link of the focal stock respectively. The dependent variables are focal stock returns in the next two to five trading days. The control variables include firm size (*Size*, taking logarithms), book-to-market ratio (*BM*), earnings-to-price ratio (*EP*), return on equity (*ROE*), past daily return (*Rtn*), total assets growth (*Assets_growth*), and daily turnover rate (*Turnover*). Non-return variables are winsorized at 1% and 99% in the cross-section and all independent variables are cross-sectionally standardized. Newey and West (1987) adjusted t-statistics are shown in parentheses. For brevity, all coefficients are shown multiplied by 100.

Table 8: The placebo portfolio sorting test of the LF and PE momentum

Panel A: Placebo LF momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.62	0.54	0.49	0.56	0.22	-0.13	-0.12	-0.14	-0.13	-0.42
	(0.77)	(0.70)	(0.64)	(0.72)	(0.29)	(-0.19)	(-0.18)	(-0.20)	(-0.19)	(-0.60)
2	1.21	1.13	1.09	1.17	0.79	0.60	0.61	0.57	0.62	0.27
	(1.53)	(1.51)	(1.46)	(1.53)	(1.04)	(0.91)	(0.95)	(0.88)	(0.95)	(0.42)
3	1.75	1.67	1.62	1.71	1.33	1.17	1.19	1.13	1.21	0.82
	(2.21)	(2.24)	(2.18)	(2.24)	(1.77)	(1.84)	(1.93)	(1.83)	(1.93)	(1.31)
4	2.11	2.02	1.98	2.07	1.71	1.96	1.99	1.95	2.02	1.71
	(2.68)	(2.75)	(2.70)	(2.75)	(2.29)	(3.01)	(3.16)	(3.06)	(3.15)	(2.63)
5	4.74	4.64	4.59	4.69	4.34	2.72	2.71	2.71	2.74	2.48
	(5.73)	(6.02)	(5.98)	(5.95)	(5.56)	(3.92)	(4.11)	(4.10)	(4.09)	(3.68)
5-1	**4.09**	**4.08**	**4.08**	**4.10**	**4.11**	**2.85**	**2.83**	**2.85**	**2.87**	**2.91**
	(12.07)	**(11.83)**	**(12.01)**	**(11.85)**	**(12.14)**	**(6.37)**	**(6.11)**	**(6.16)**	**(6.20)**	**(6.48)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Placebo PE momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.65	0.57	0.53	0.60	0.26	0.10	0.13	0.10	0.12	-0.21
	(0.81)	(0.75)	(0.70)	(0.77)	(0.33)	(0.14)	(0.19)	(0.15)	(0.17)	(-0.30)
2	1.35	1.26	1.21	1.29	0.93	0.64	0.66	0.62	0.67	0.35
	(1.72)	(1.69)	(1.64)	(1.71)	(1.25)	(1.01)	(1.05)	(0.99)	(1.06)	(0.55)
3	1.63	1.55	1.50	1.60	1.20	1.10	1.12	1.08	1.15	0.83
	(2.09)	(2.11)	(2.04)	(2.13)	(1.61)	(1.73)	(1.79)	(1.70)	(1.81)	(1.28)
4	2.14	2.05	2.01	2.10	1.74	1.83	1.85	1.82	1.87	1.55
	(2.70)	(2.78)	(2.73)	(2.78)	(2.32)	(2.83)	(2.98)	(2.92)	(2.98)	(2.45)
5	4.61	4.52	4.47	4.57	4.23	2.65	2.68	2.67	2.69	2.48
	(5.54)	(5.82)	(5.79)	(5.76)	(5.38)	(3.88)	(4.08)	(4.03)	(4.05)	(3.69)
5-1	**3.94**	**3.92**	**3.91**	**3.95**	**3.96**	**2.55**	**2.55**	**2.56**	**2.57**	**2.69**
	(10.86)	**(10.63)**	**(10.82)**	**(10.65)**	**(10.89)**	**(5.77)**	**(5.55)**	**(5.59)**	**(5.62)**	**(5.93)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of Placebo LF and PE momentum. At the end of each trading day, rather than defining LF links according to where stocks are mentioned, we randomly define 50% of CM links as LF links (denoted as Placebo LF), then the other 50% as PE links (denoted as Placebo PE). The predictive signal of one focal stock is the weighted average return of its connected stocks defined by the placebo link. For each placebo momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. We report the mean return of each portfolio, as well as returns adjusted by the Fama and French (1993) three-factor model, Fama and French (2015) five-factor model, Carhart (1997) four-factor model, and Liu et al. (2019) Chinese four-factor model. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 9: Placebo portfolio return spreads

Panel A: Equal-weighted		
	Long-only difference	Long-short difference
Placebo PE minus Placebo LF	-0.12	-0.15
	(-0.92)	(-0.80)
Placebo PE minus CM	-0.25	-0.30
	(-1.87)	(-1.75)
Placebo LF minus CM	-0.13	-0.15
	(-1.68)	(-1.43)
Panel B: Value-weighted		
	Long-only difference	Long-short difference
Placebo PE minus Placebo LF	-0.07	-0.29
	(-0.32)	(-0.96)
Placebo PE minus CM	-0.38	-0.83
	(-1.93)	**(-3.25)**
Placebo LF minus CM	**-0.31**	**-0.54**
	(-2.42)	**(-2.98)**

The table shows statistical test results of long-only and long-short portfolio return spreads between the Placebo PE momentum, Placebo LF momentum, and CM momentum. The return spread is computed by taking the difference between the time series returns of two portfolios. All daily returns and alphas are converted into monthly percentages using compound interest. The sample period is 2012 - 2021. Newey and West (1987) adjusted t-statistics are shown in parentheses. Return spreads with t-statistics higher than 2.00 are highlighted in bold.

Table 10: Fundamental linkages

Panel A: Dependent variable: *Sales growth* (t)	1	2	3	4	5	6
LF sales growth (t-1)	0.716		1.106			
	(1.36)		(1.64)			
PE sales growth (t-1)		0.830	0.972			
		(1.48)	(1.39)			
LF sales growth (t)				2.482		2.476
				(4.09)		(2.87)
PE sales growth (t)					3.146	3.741
					(3.89)	(4.46)
Control	YES	YES	YES	YES	YES	YES
Entity effects	YES	YES	YES	YES	YES	YES
Time effects	YES	YES	YES	YES	YES	YES
Adj. R Square	0.006	0.005	0.008	0.010	0.017	0.022
Total Obs.	6317	6228	4457	7738	7925	5621
Avg. Obs.	790	779	557	860	881	625
Panel B: Dependent variable: *Profit growth* (t)						
	1	2	3	4	5	6
LF profit growth (t-1)	0.025		0.065			
	(0.21)		(0.49)			
PE profit growth (t-1)		0.023	-0.017			
		(0.25)	(-0.17)			
LF profit growth (t)				0.443		0.406
				(6.18)		(6.25)
PE profit growth (t)					0.800	0.859
					(3.37)	(3.41)
Control	YES	YES	YES	YES	YES	YES
Entity effects	YES	YES	YES	YES	YES	YES
Time effects	YES	YES	YES	YES	YES	YES
Adj. R Square	0.008	0.008	0.009	0.017	0.026	0.038
Total Obs.	7711	7565	5624	9239	9374	6896
Avg. Obs.	857	841	625	924	937	690

This table shows the results of fundamentals panel regressions following Ali and Hirshleifer (2020). In Panel A, the dependent variable is the annual sales growth, computed as the percent change of *Sales* per share at time t to *Sales* per share at time $t-1$; while in Panel B, the dependent variable is the annual profit growth, computed as the difference between $Profit$ at time t and $Profit$ at time $t-1$, divided by the average of total assets at time t and $t-1$. *LF sales growth* is calculated as the weighted average *Sales growth* of leader stocks of the focal stock; while *PE sales growth* is calculated as the weighted average *Sales growth* of peer stocks of the focal stock. Profit growth measures are calculated similarly. The sample contains firms with December fiscal year ends. Control variables include firm size and book-to-market ratio. The sample period is 2012 - 2021. All variables are based at the end of each calendar year and are winsorized at the 1% and 99% levels. Independent variables are cross-sectionally standardized. All regressions include entity-fixed and time-fixed effects. T-statistics based on standard errors clustered by entity and time are shown in parentheses.

Table 11: The big-leader and small-leader LF momentum

Panel A: The big-leader LF momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	1.06	0.99	0.94	1.01	0.65	0.18	0.22	0.20	0.20	-0.09
	(1.34)	(1.31)	(1.26)	(1.32)	(0.86)	(0.27)	(0.32)	(0.29)	(0.29)	(-0.14)
2	1.45	1.37	1.33	1.40	1.03	0.81	0.82	0.80	0.83	0.51
	(1.87)	(1.86)	(1.81)	(1.87)	(1.39)	(1.26)	(1.28)	(1.25)	(1.28)	(0.79)
3	1.55	1.48	1.43	1.52	1.14	1.05	1.10	1.05	1.11	0.75
	(2.05)	(2.06)	(2.00)	(2.08)	(1.57)	(1.72)	(1.82)	(1.71)	(1.83)	(1.22)
4	1.72	1.65	1.63	1.69	1.35	1.55	1.59	1.58	1.60	1.36
	(2.27)	(2.32)	(2.29)	(2.33)	(1.87)	(2.48)	(2.64)	(2.60)	(2.64)	(2.19)
5	2.34	2.25	2.23	2.29	1.98	1.93	1.92	1.91	1.95	1.70
	(2.96)	(3.04)	(3.03)	(3.03)	(2.65)	(2.84)	(2.98)	(2.94)	(2.97)	(2.58)
5-1	**1.27**	**1.25**	**1.27**	**1.27**	**1.32**	**1.74**	**1.70**	**1.71**	**1.75**	**1.80**
	(5.31)	**(5.18)**	**(5.23)**	**(5.23)**	**(5.53)**	**(4.24)**	**(4.03)**	**(4.05)**	**(4.11)**	**(4.46)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: The small-leader LF momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	1.13	1.04	0.98	1.08	0.71	0.29	0.31	0.30	0.31	0.03
	(1.39)	(1.35)	(1.28)	(1.38)	(0.90)	(0.45)	(0.50)	(0.47)	(0.49)	(0.05)
2	1.33	1.25	1.21	1.29	0.93	0.95	0.97	0.99	0.98	0.72
	(1.68)	(1.67)	(1.62)	(1.69)	(1.23)	(1.51)	(1.57)	(1.58)	(1.58)	(1.14)
3	1.75	1.68	1.65	1.72	1.32	1.19	1.24	1.23	1.23	0.96
	(2.23)	(2.26)	(2.24)	(2.27)	(1.77)	(1.96)	(2.06)	(2.02)	(2.05)	(1.56)
4	1.74	1.65	1.61	1.70	1.34	0.82	0.88	0.83	0.89	0.61
	(2.21)	(2.22)	(2.17)	(2.24)	(1.77)	(1.36)	(1.45)	(1.35)	(1.47)	(0.98)
5	5.04	4.96	4.90	5.02	4.63	1.96	2.02	2.02	2.03	1.76
	(5.64)	(5.86)	(5.79)	(5.84)	(5.43)	(3.06)	(3.19)	(3.16)	(3.19)	(2.75)
5-1	**3.88**	**3.88**	**3.88**	**3.90**	**3.90**	**1.67**	**1.70**	**1.72**	**1.72**	**1.72**
	(8.41)	**(8.46)**	**(8.54)**	**(8.47)**	**(8.52)**	**(4.51)**	**(4.50)**	**(4.49)**	**(4.58)**	**(4.62)**
SpearmanR	0.90	0.90	0.90	0.90	1.00	0.70	0.70	0.70	0.70	0.70
P-value	0.04	0.04	0.04	0.04	0.00	0.19	0.19	0.19	0.19	0.19

This table reports the portfolio sorting results of the decomposition of the LF momentum according to the firm size. We divide the LF momentum into two parts: the big-leader LF momentum and the small-leader LF momentum. For the big-leader LF momentum, the predictive signal LF_Rtn only considers leader stocks with market capitalization in the top 50% of the total sample, while for the small-leader LF momentum, only leader stocks with market capitalization in the bottom 50% of the total sample are included. Then we do group portfolio sorting according to the two kinds of LF_Rtn. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. We report the mean return of each portfolio as well as returns adjusted by the Fama and French (1993) three-factor model, Fama and French (2015) five-factor model, Carhart (1997) four-factor model, and Liu et al. (2019) Chinese four-factor model. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 12: The LF and PE momentum after removing LF-only links

Panel A: LF2 momentum (removing LF-only links)										
	Equal-weighte					Value-weighte				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.83	0.77	0.71	0.79	0.44	0.39	0.45	0.41	0.44	0.15
	(1.05)	(1.02)	(0.95)	(1.03)	(0.58)	(0.58)	(0.67)	(0.62)	(0.65)	(0.22)
2	1.19	1.12	1.08	1.15	0.78	0.70	0.74	0.70	0.74	0.41
	(1.54)	(1.52)	(1.47)	(1.54)	(1.05)	(1.11)	(1.18)	(1.12)	(1.17)	(0.64)
3	1.42	1.35	1.31	1.39	1.02	1.11	1.13	1.10	1.15	0.82
	(1.87)	(1.88)	(1.83)	(1.90)	(1.41)	(1.81)	(1.87)	(1.81)	(1.89)	(1.32)
4	1.92	1.85	1.82	1.90	1.54	1.54	1.58	1.55	1.59	1.31
	(2.52)	(2.59)	(2.55)	(2.60)	(2.14)	(2.46)	(2.60)	(2.53)	(2.60)	(2.10)
5	4.41	4.33	4.29	4.37	4.05	2.17	2.18	2.17	2.20	1.97
	(5.38)	(5.62)	(5.58)	(5.56)	(5.20)	(3.07)	(3.23)	(3.23)	(3.22)	(2.88)
5-1	**3.56**	**3.53**	**3.55**	**3.56**	**3.59**	**1.77**	**1.72**	**1.75**	**1.76**	**1.81**
	(9.57)	**(9.41)**	**(9.55)**	**(9.44)**	**(9.65)**	**(4.02)**	**(3.85)**	**(3.85)**	**(3.91)**	**(4.08)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: PE2 momentum (removing LF-only links)										
	Equal-weighte					Value-weighte				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.23	0.15	0.10	0.17	-0.18	-0.43	-0.41	-0.45	-0.43	-0.72
	(0.28)	(0.19)	(0.13)	(0.21)	(-0.23)	(-0.62)	(-0.60)	(-0.65)	(-0.62)	(-1.05)
2	1.09	1.00	0.95	1.03	0.66	0.70	0.71	0.65	0.72	0.36
	(1.36)	(1.32)	(1.27)	(1.34)	(0.86)	(1.05)	(1.09)	(1.00)	(1.09)	(0.54)
3	1.81	1.72	1.67	1.76	1.38	1.13	1.14	1.10	1.16	0.84
	(2.29)	(2.32)	(2.26)	(2.33)	(1.84)	(1.77)	(1.83)	(1.76)	(1.83)	(1.31)
4	2.15	2.07	2.04	2.12	1.75	1.90	1.93	1.89	1.96	1.65
	(2.71)	(2.79)	(2.75)	(2.79)	(2.32)	(2.96)	(3.10)	(3.01)	(3.11)	(2.57)
5	5.34	5.24	5.19	5.29	4.95	3.02	3.03	3.01	3.06	2.80
	(6.33)	(6.67)	(6.64)	(6.60)	(6.23)	(4.27)	(4.46)	(4.43)	(4.45)	(4.03)
5-1	**5.10**	**5.08**	**5.09**	**5.11**	**5.14**	**3.46**	**3.46**	**3.47**	**3.50**	**3.54**
	(12.89)	**(12.57)**	**(12.75)**	**(12.62)**	**(12.95)**	**(7.09)**	**(6.80)**	**(6.91)**	**(6.91)**	**(7.24)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the LF2 and PE2 momentum after removing those links that have only been identified as LF links (LF-only links). The CM link considers stocks appearing in the same sentence of a news article during the previous 90 days as news co-mention connected stocks. The LF2 link defines a directional relationship from the stock appearing in the news title (leader) to the stock mentioned in the body text of the same news (follower) during the previous 90 days, excluding those that have never been identified as CM links. The PE2 link is constructed by subtracting all LF2 links from CM links. The predictive signal of one focal stock is the average return of its connected stocks defined by each of the three links, weighted by the co-mention times (for LF2 momentum, the weight is the leading times). For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. We report the mean return of each portfolio as well as returns adjusted by the Fama and French (1993) three-factor model, Fama and French (2015) five-factor model, Carhart (1997) four-factor model, and Liu et al. (2019) Chinese four-factor model. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 13: Portfolio return spreads after removing LF-only links

Panel A: Equal-weighted		
	Long-only difference	Long-short difference
PE2 minus LF2	**0.89**	**1.50**
	(4.16)	**(5.78)**
PE2 minus CM	**0.44**	**0.82**
	(3.97)	**(5.91)**
LF2 minus CM	**-0.44**	**-0.66**
	(-2.40)	**(-3.18)**
Panel B: Value-weighted		
	Long-only difference	Long-short difference
PE2 minus LF2	**0.83**	**1.66**
	(3.06)	**(4.29)**
PE2 minus CM	-0.02	0.04
	(-0.13)	(0.19)
LF2 minus CM	**-0.84**	**-1.59**
	(-3.94)	**(-5.25)**

This table shows statistical test results of long-only and long-short portfolio return spreads between the PE2, LF2, and CM momentum after removing those links that have only been identified as LF links (LF-only links). The return spread is computed by taking the difference between the time series returns of two portfolios. All daily returns and alphas are converted into monthly percentages using compound interest. The sample period is 2012 - 2021. Newey and West (1987) adjusted t-statistics are shown in parentheses. Return spreads with t-statistics higher than 2.00 are highlighted in bold.

Table 14: Fama-Macbeth regressions after removing LF-only links

	1	2	3	4	5	6	7
LF2_Rtn	0.0120			0.0054		0.0001	0.0042
	(2.85)			(1.15)		(0.02)	(0.67)
PE2_Rtn		0.0313		0.0160	0.0223		0.0150
		(6.93)		(2.64)	(3.84)		(2.04)
CM_Rtn			0.0240		0.0101	0.0167	0.0020
			(6.83)		(1.78)	(2.61)	(0.19)
Size	-0.0794	-0.0858	-0.0922	-0.0729	-0.0854	-0.0783	-0.0724
	(-9.70)	(-10.73)	(-11.72)	(-8.74)	(-10.68)	(-9.60)	(-9.16)
BM	0.0082	0.0120	0.0094	0.0138	0.0120	0.0084	0.0136
	(1.30)	(1.94)	(1.59)	(2.10)	(1.94)	(1.34)	(2.16)
EP	0.0030	0.0024	0.0037	-0.0016	0.0023	0.0024	-0.0018
	(0.65)	(0.52)	(0.85)	(-0.32)	(0.49)	(0.53)	(-0.38)
ROE	0.0410	0.0467	0.0456	0.0465	0.0467	0.0409	0.0458
	(7.05)	(8.77)	(9.10)	(6.96)	(8.77)	(7.06)	(7.03)
Rtn	0.2036	0.1968	0.2097	0.1800	0.1955	0.2005	0.1784
	(17.81)	(18.33)	(19.90)	(15.72)	(18.27)	(17.97)	(16.94)
Assets_growth	0.0423	0.0363	0.0371	0.0370	0.0359	0.0418	0.0368
	(9.71)	(9.83)	(10.63)	(7.85)	(9.78)	(9.66)	(7.99)
Turnover	-0.1470	-0.1475	-0.1567	-0.1335	-0.1475	-0.1470	-0.1328
	(-14.30)	(-15.49)	(-17.10)	(-12.43)	(-15.49)	(-14.34)	(-12.42)
Intercept	0.1050	0.1027	0.1013	0.1044	0.1027	0.1044	0.1060
	(2.55)	(2.50)	(2.48)	(2.52)	(2.49)	(2.54)	(2.60)
Avg. R Square	0.1061	0.0968	0.0885	0.1225	0.0976	0.1083	0.1241
Total Obs.	1247865	1965781	2597936	1029250	1965781	1247865	1029250
Avg. Obs.	513	809	1069	423	809	513	423

This table reports the results of Fama and MacBeth (1973) regressions after removing those links that have only been identified as LF links (LF-only links). The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. The dependent variable is the focal stock return in the next trading day; while independent variables are $LF2\_Rtn$, $PE2\_Rtn$, and CM_Rtn which are the connected stock return by the LF2, PE2, and CM link of the focal stock respectively. The control variables include firm size (*Size*, taking logarithms), book-to-market ratio (BM), earnings-to-price ratio (EP), return on equity (ROE), past daily return (*Rtn*), total assets growth (*Assets_growth*), and daily turnover rate ($Turnover$). Non-return variables are winsorized at 1% and 99% in the cross-section and all independent variables are cross-sectionally standardized. Newey and West (1987) adjusted t-statistics are shown in parentheses. For brevity, all coefficients are shown multiplied by 100.

Table 15: The LF, PE, CM momentum with transaction costs

Panel A: LF momentum with transaction cost										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	-1.19	-1.27	-1.32	-1.24	-1.61	-1.99	-1.97	-1.98	-1.98	-2.26
	(-1.50)	(-1.68)	(-1.75)	(-1.62)	(-2.10)	(-2.94)	(-2.94)	(-2.96)	(-2.95)	(-3.34)
2	-0.84	-0.92	-0.97	-0.89	-1.26	-1.49	-1.47	-1.50	-1.46	-1.78
	(-1.09)	(-1.26)	(-1.33)	(-1.19)	(-1.71)	(-2.36)	(-2.33)	(-2.40)	(-2.30)	(-2.81)
3	-0.65	-0.73	-0.77	-0.69	-1.06	-1.12	-1.11	-1.15	-1.09	-1.45
	(-0.86)	(-1.01)	(-1.08)	(-0.93)	(-1.45)	(-1.86)	(-1.89)	(-1.95)	(-1.83)	(-2.41)
4	-0.35	-0.43	-0.46	-0.39	-0.73	-0.59	-0.57	-0.59	-0.55	-0.81
	(-0.45)	(-0.60)	(-0.64)	(-0.53)	(-1.00)	(-0.95)	(-0.95)	(-0.97)	(-0.91)	(-1.31)
5	1.44	1.34	1.31	1.39	1.06	-0.15	-0.15	-0.17	-0.12	-0.38
	(1.77)	(1.77)	(1.74)	(1.80)	(1.38)	(-0.21)	(-0.22)	(-0.25)	(-0.18)	(-0.57)
5-1	0.42	0.41	0.43	0.43	0.47	-0.34	-0.36	-0.36	-0.32	-0.30
	(1.56)	(1.49)	(1.56)	(1.55)	(1.71)	(-0.82)	(-0.86)	(-0.86)	(-0.75)	(-0.74)
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: PE momentum with transaction cost										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	-1.95	-2.03	-2.08	-2.01	-2.35	-2.59	-2.58	-2.61	-2.60	-2.88
	(-2.42)	(-2.62)	(-2.70)	(-2.57)	(-3.02)	(-3.81)	(-3.80)	(-3.86)	(-3.81)	(-4.25)
2	-1.11	-1.20	-1.24	-1.16	-1.53	-1.49	-1.48	-1.54	-1.47	-1.83
	(-1.41)	(-1.61)	(-1.67)	(-1.53)	(-2.04)	(-2.27)	(-2.30)	(-2.40)	(-2.26)	(-2.82)
3	-0.41	-0.50	-0.54	-0.46	-0.83	-1.06	-1.06	-1.10	-1.04	-1.36
	(-0.53)	(-0.68)	(-0.74)	(-0.61)	(-1.11)	(-1.68)	(-1.72)	(-1.77)	(-1.66)	(-2.15)
4	-0.07	-0.15	-0.18	-0.10	-0.47	-0.32	-0.29	-0.32	-0.26	-0.56
	(-0.09)	(-0.20)	(-0.25)	(-0.14)	(-0.62)	(-0.50)	(-0.47)	(-0.52)	(-0.42)	(-0.89)
5	3.06	2.96	2.91	3.00	2.68	0.78	0.79	0.78	0.82	0.56
	(3.66)	(3.81)	(3.76)	(3.79)	(3.40)	(1.12)	(1.18)	(1.15)	(1.20)	(0.82)
5-1	**2.82**	**2.81**	**2.81**	**2.83**	**2.86**	**1.21**	**1.21**	**1.22**	**1.25**	**1.29**
	(7.21)	**(7.01)**	**(7.12)**	**(7.07)**	**(7.29)**	**(2.51)**	**(2.41)**	**(2.46)**	**(2.49)**	**(2.67)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: CM momentum with transaction cost										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	-1.58	-1.67	-1.72	-1.64	-1.98	-2.53	-2.52	-2.55	-2.53	-2.83
	(-1.95)	(-2.17)	(-2.25)	(-2.11)	(-2.56)	(-3.63)	(-3.64)	(-3.70)	(-3.64)	(-4.05)
2	-1.08	-1.16	-1.20	-1.13	-1.51	-1.84	-1.85	-1.89	-1.84	-2.19
	(-1.36)	(-1.55)	(-1.62)	(-1.47)	(-2.00)	(-2.76)	(-2.83)	(-2.92)	(-2.78)	(-3.32)
3	-0.52	-0.61	-0.65	-0.56	-0.93	-0.81	-0.80	-0.85	-0.77	-1.15
	(-0.67)	(-0.83)	(-0.89)	(-0.76)	(-1.27)	(-1.26)	(-1.27)	(-1.36)	(-1.20)	(-1.80)
4	-0.15	-0.24	-0.28	-0.19	-0.55	-0.31	-0.30	-0.33	-0.27	-0.57
	(-0.19)	(-0.32)	(-0.38)	(-0.25)	(-0.73)	(-0.47)	(-0.47)	(-0.52)	(-0.42)	(-0.88)
5	2.60	2.50	2.45	2.54	2.20	0.80	0.78	0.77	0.81	0.55
	(3.16)	(3.26)	(3.21)	(3.25)	(2.84)	(1.15)	(1.19)	(1.17)	(1.21)	(0.82)
5-1	**1.98**	**1.97**	**1.97**	**1.99**	**2.00**	**1.17**	**1.14**	**1.16**	**1.17**	**1.22**
	(6.30)	**(6.14)**	**(6.23)**	**(6.18)**	**(6.35)**	**(2.56)**	**(2.42)**	**(2.46)**	**(2.50)**	**(2.66)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of the LF, PE, and CM momentum after considering a transaction cost of 11 bps (buy and sell combined), respectively. The sample stocks include all listed stocks on the main board of the Shanghai Stock Exchange, Shenzhen Stock Exchange, and Growth Enterprise Market (GEM). ST shares are excluded. The sample period starts from January 2012 to December 2021. For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. We report the mean return of each portfolio as well as returns adjusted by the Fama and French (1993) three-factor model, Fama and French (2015) five-factor model, Carhart (1997) four-factor model, and Liu et al. (2019) Chinese four-factor modal. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 16: LF, PE, and CM momentum under different identification windows

Panel A: Mean returns									
	LF momentum			PE momentum			CM momentum		
	30-D	90-D	180-D	30-D	90-D	180-D	30-D	90-D	180-D
1	0.79	1.01	1.01	0.11	0.23	0.33	0.50	0.61	0.58
	(1.02)	(1.25)	(1.25)	(0.14)	(0.28)	(0.40)	(0.62)	(0.74)	(0.71)
2	1.35	1.37	1.28	1.02	1.09	1.10	1.27	1.13	1.15
	(1.78)	(1.75)	(1.61)	(1.32)	(1.36)	(1.36)	(1.65)	(1.41)	(1.43)
3	1.74	1.56	1.61	1.90	1.80	1.80	1.72	1.70	1.72
	(2.34)	(2.02)	(2.04)	(2.51)	(2.29)	(2.26)	(2.27)	(2.16)	(2.14)
4	1.75	1.87	2.01	2.13	2.15	2.27	2.02	2.07	2.23
	(2.34)	(2.41)	(2.55)	(2.76)	(2.71)	(2.83)	(2.61)	(2.60)	(2.77)
5	4.88	3.69	3.26	7.66	5.34	4.44	6.73	4.88	4.18
	(5.82)	(4.50)	(4.01)	(8.18)	(6.33)	(5.36)	(7.65)	(5.86)	(5.07)
5-1	4.06	2.66	2.22	7.54	5.11	4.10	6.20	4.25	3.58
	(10.11)	(9.62)	(9.64)	(12.84)	(12.91)	(13.01)	(14.10)	(13.36)	(12.96)
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: CH-4 alphas									
	LF momentum			PE momentum			CM momentum		
	30-D	90-D	180-D	30-D	90-D	180-D	30-D	90-D	180-D
1	0.42	0.58	0.58	-0.26	-0.18	-0.09	0.12	0.20	0.17
	(0.56)	(0.75)	(0.75)	(-0.34)	(-0.23)	(-0.11)	(0.16)	(0.25)	(0.21)
2	0.95	0.94	0.85	0.61	0.66	0.66	0.85	0.68	0.71
	(1.29)	(1.25)	(1.12)	(0.82)	(0.87)	(0.85)	(1.15)	(0.89)	(0.92)
3	1.34	1.15	1.17	1.48	1.38	1.36	1.29	1.27	1.28
	(1.86)	(1.56)	(1.57)	(2.02)	(1.84)	(1.80)	(1.76)	(1.70)	(1.68)
4	1.39	1.48	1.60	1.76	1.75	1.85	1.64	1.66	1.81
	(1.95)	(2.01)	(2.14)	(2.38)	(2.31)	(2.44)	(2.21)	(2.20)	(2.38)
5	4.53	3.30	2.86	7.33	4.96	4.04	6.36	4.47	3.78
	(5.64)	(4.27)	(3.74)	(8.14)	(6.23)	(5.20)	(7.60)	(5.71)	(4.89)
5-1	4.09	2.71	2.26	7.61	5.14	4.13	6.23	4.27	3.60
	(10.21)	(9.71)	(9.60)	(12.94)	(12.97)	(12.98)	(14.11)	(13.39)	(12.90)
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of the LF, PE, and CM momentum under three identification windows including 30-day, 90-day, and 180-day, respectively. The LF link defines a directional relationship from the stocks appearing in the news title (leaders) to the stocks mentioned in the body text of the same news (followers) during the identification window. The PE link defines two firms to be peer stocks with each other if they are only mentioned in the same news body during the identification window. The CM link considers stocks appearing in the same sentence of a news article during the identification window as news co-mention connected stocks. The predictive signal of one focal stock is the average return of its connected stocks defined by each of the three links, weighted by the co-mention times (for LF momentum, the weight is the leading times). For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. Panel A reports the mean returns of portfolios, while Panel B reports the Liu et al. (2019) Chinese four-factor adjusted alphas. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 17: Portfolio return differences under different identification windows

Panel A: PE minus LF						
	Long-only difference			Long-short difference		
News windows	30-D	90-D	180-D	30-D	90-D	180-D
Return Diff	**2.65**	**1.60**	**1.15**	**3.35**	**2.39**	**1.84**
T-statistics	**(7.51)**	**(8.03)**	**(8.38)**	**(7.80)**	**(9.05)**	**(9.41)**
Panel B: PE minus CM						
	Long-only difference			Long-short difference		
News windows	30-D	90-D	180-D	30-D	90-D	180-D
Return Diff	**0.87**	**0.44**	**0.25**	**1.27**	**0.83**	**0.50**
T-statistics	**(4.36)**	**(3.99)**	**(3.39)**	**(5.41)**	**(5.94)**	**(5.12)**
Panel C: LF minus CM						
	Long-only difference			Long-short difference		
News windows	30-D	90-D	180-D	30-D	90-D	180-D
Return Diff	**-1.73**	**-1.14**	**-0.89**	**-2.02**	**-1.53**	**-1.31**
T-statistics	**(-7.38)**	**(-7.61)**	**(-7.97)**	**(-6.89)**	**(-7.95)**	**(-8.58)**

The table shows the statistical test results of long-only and long-short portfolio return spreads between the PE, LF, and CM momentum under three identification windows including 30-day, 90-day, and 180-day respectively. The return spread is computed by taking the difference between the time series returns of two portfolios. All daily returns and alphas are converted into monthly percentages using compound interest. The sample period is 2012 - 2021. Newey and West (1987) adjusted t-statistics are shown in parentheses. Return spreads with t-statistics higher than 2.00 are highlighted in bold.

Table 18: The LF, PE, and CM momentum under the *same_article* co-mention type

Panel A: LF momentum										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	1.01	0.93	0.87	0.95	0.58	0.19	0.22	0.20	0.20	-0.09
	(1.25)	(1.21)	(1.14)	(1.23)	(0.75)	(0.29)	(0.32)	(0.29)	(0.29)	(-0.13)
2	1.37	1.28	1.24	1.32	0.94	0.71	0.72	0.69	0.73	0.40
	(1.75)	(1.73)	(1.68)	(1.75)	(1.25)	(1.11)	(1.14)	(1.10)	(1.14)	(0.63)
3	1.56	1.48	1.44	1.52	1.15	1.08	1.08	1.04	1.11	0.75
	(2.02)	(2.03)	(1.98)	(2.05)	(1.56)	(1.76)	(1.82)	(1.75)	(1.84)	(1.23)
4	1.87	1.78	1.75	1.83	1.48	1.62	1.64	1.62	1.66	1.40
	(2.41)	(2.46)	(2.42)	(2.47)	(2.01)	(2.58)	(2.72)	(2.65)	(2.72)	(2.24)
5	3.69	3.60	3.56	3.64	3.30	2.07	2.07	2.05	2.10	1.83
	(4.50)	(4.70)	(4.68)	(4.66)	(4.27)	(2.95)	(3.09)	(3.07)	(3.08)	(2.69)
5-1	**2.66**	**2.65**	**2.66**	**2.66**	**2.71**	**1.88**	**1.85**	**1.85**	**1.90**	**1.92**
	(9.62)	**(9.42)**	**(9.50)**	**(9.46)**	**(9.71)**	**(4.49)**	**(4.35)**	**(4.36)**	**(4.42)**	**(4.72)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: PE momentum (*same_article* type)										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.58	0.49	0.44	0.51	0.17	-0.34	-0.34	-0.38	-0.35	-0.63
	(0.71)	(0.63)	(0.57)	(0.65)	(0.21)	(-0.49)	(-0.50)	(-0.55)	(-0.51)	(-0.92)
2	1.15	1.06	1.00	1.10	0.69	0.69	0.71	0.65	0.71	0.33
	(1.43)	(1.39)	(1.33)	(1.42)	(0.90)	(1.05)	(1.09)	(1.00)	(1.09)	(0.50)
3	1.78	1.69	1.65	1.73	1.37	1.31	1.32	1.27	1.34	1.00
	(2.24)	(2.26)	(2.21)	(2.27)	(1.80)	(2.00)	(2.07)	(1.99)	(2.08)	(1.54)
4	2.11	2.02	1.99	2.07	1.71	1.83	1.84	1.82	1.88	1.58
	(2.65)	(2.71)	(2.68)	(2.71)	(2.25)	(2.72)	(2.85)	(2.80)	(2.86)	(2.37)
5	4.73	4.62	4.57	4.67	4.33	2.77	2.77	2.74	2.80	2.52
	(5.63)	(5.94)	(5.90)	(5.87)	(5.49)	(3.97)	(4.15)	(4.09)	(4.12)	(3.69)
5-1	**4.13**	**4.11**	**4.11**	**4.13**	**4.15**	**3.12**	**3.12**	**3.12**	**3.16**	**3.17**
	(12.80)	**(12.53)**	**(12.67)**	**(12.54)**	**(12.90)**	**(6.93)**	**(6.66)**	**(6.80)**	**(6.72)**	**(7.14)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: CM momentum (*same_article* type)										
	Equal-weighted					Value-weighted				
	Mean	FF-3	FF-5	Carhart-4	CH-4	Mean	FF-3	FF-5	Carhart-4	CH-4
1	0.81	0.71	0.66	0.74	0.39	-0.30	-0.31	-0.34	-0.31	-0.60
	(0.98)	(0.91)	(0.86)	(0.93)	(0.50)	(-0.43)	(-0.45)	(-0.49)	(-0.45)	(-0.87)
2	1.24	1.15	1.09	1.19	0.80	0.45	0.47	0.43	0.47	0.12
	(1.54)	(1.51)	(1.44)	(1.54)	(1.04)	(0.67)	(0.71)	(0.65)	(0.70)	(0.18)
3	1.66	1.57	1.53	1.61	1.23	1.08	1.06	1.03	1.09	0.71
	(2.08)	(2.08)	(2.04)	(2.10)	(1.62)	(1.62)	(1.65)	(1.60)	(1.66)	(1.09)
4	2.16	2.07	2.02	2.12	1.75	2.04	2.05	1.99	2.08	1.75
	(2.71)	(2.77)	(2.72)	(2.77)	(2.31)	(3.04)	(3.16)	(3.05)	(3.16)	(2.61)
5	4.48	4.36	4.32	4.41	4.06	2.97	2.95	2.92	2.97	2.72
	(5.37)	(5.65)	(5.61)	(5.59)	(5.20)	(4.15)	(4.38)	(4.34)	(4.35)	(3.97)
5-1	**3.65**	**3.63**	**3.63**	**3.65**	**3.66**	**3.28**	**3.27**	**3.27**	**3.30**	**3.34**
	(13.20)	**(12.87)**	**(13.07)**	**(12.87)**	**(13.35)**	**(7.25)**	**(6.95)**	**(7.10)**	**(6.99)**	**(7.52)**
SpearmanR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of the LF, PE, and CM momentum after adjusting the identification methodology of the CM link to *same_article* strategy. Here, the CM link defines two stocks to be news co-mention connected as long as they appear in the same news article during the 90-day identification window. The LF link defines a directional relationship from the stock appearing in the news title (leader) to the stock mentioned in the body text of the same news (follower) during the identification window. The PE link is constructed by subtracting all LF links from CM links. The predictive signal of one focal stock is the average return of its connected stocks defined by each of the three links, weighted by the co-mention times (for LF momentum, the weight is the leading times). For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one day and are rebalanced daily. Panel A reports the mean returns of portfolios, while Panel B reports the Liu et al. (2019) Chinese four-factor adjusted alphas. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Table 19: Portfolio return differences under the *same_article* co-mention type

Panel A: Equal-weighted		
	Long-only difference	Long-short difference
PE minus LF	**1.01**	**1.44**
	(5.98)	**(6.57)**
PE minus CM	**0.24**	**0.47**
	(2.59)	**(3.88)**
LF minus CM	**-0.76**	**-0.96**
	(-5.32)	**(-5.51)**
Panel B: Value-weighted		
	Long-only difference	Long-short difference
PE minus LF	**0.68**	**1.22**
	(3.04)	**(3.96)**
PE minus CM	-0.20	-0.16
	(-1.36)	(-0.80)
LF minus CM	**-0.87**	**-1.37**
	(-4.41)	**(-5.05)**

The table shows the statistical test results of long-only and long-short portfolio return spreads between the PE, LF, and CM momentum after adjusting the identification methodology of the CM link to *same_article* strategy. Here, the CM link defines two stocks to be news co-mention connected as long as they appear in the same news article during the 90-day identification window. The return spread is computed by taking the difference between the time series returns of two portfolios. All daily returns and alphas are converted into monthly percentages using compound interest. The sample period is 2012 - 2021. Newey and West (1987) adjusted t-statistics are shown in parentheses. Return spreads with t-statistics higher than 2.00 are highlighted in bold.

Table 20: The weekly performance of the LF, PE, and CM momentum

Panel A: Mean return									
	LF momentum			PE momentum			CM momentum		
	30-D	90-D	180-D	30-D	90-D	180-D	30-D	90-D	180-D
1	1.21	1.26	1.26	0.83	0.91	0.95	0.83	0.96	1.00
	(1.62)	(1.66)	(1.63)	(1.13)	(1.21)	(1.26)	(1.11)	(1.25)	(1.29)
2	1.37	1.26	1.28	1.31	1.21	1.26	1.36	1.24	1.26
	(1.97)	(1.75)	(1.76)	(1.82)	(1.66)	(1.69)	(1.88)	(1.70)	(1.72)
3	1.45	1.30	1.33	1.44	1.44	1.42	1.45	1.36	1.44
	(2.10)	(1.85)	(1.84)	(1.99)	(1.96)	(1.94)	(2.05)	(1.86)	(1.94)
4	1.29	1.40	1.43	1.49	1.72	1.74	1.37	1.60	1.65
	(1.84)	(1.92)	(1.95)	(2.09)	(2.29)	(2.29)	(1.90)	(2.10)	(2.17)
5	2.62	2.42	2.25	3.86	2.92	2.60	3.42	2.82	2.55
	(3.17)	(3.01)	(2.85)	(3.90)	(3.41)	(3.17)	(3.75)	(3.42)	(3.15)
5-1	**1.40**	**1.15**	**0.98**	**3.01**	**1.99**	**1.64**	**2.57**	**1.85**	**1.54**
	(3.29)	**(4.00)**	**(4.24)**	**(4.63)**	**(4.65)**	**(4.90)**	**(5.27)**	**(5.49)**	**(5.28)**
SpearmanR	0.70	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00
P-value	0.19	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00
Panel B: CH-4 alphas									
	LF momentum			PE momentum			CM momentum		
	30-D	90-D	180-D	30-D	90-D	180-D	30-D	90-D	180-D
1	0.87	0.89	0.89	0.57	0.59	0.61	0.51	0.60	0.62
	(1.18)	(1.18)	(1.15)	(0.75)	(0.77)	(0.79)	(0.68)	(0.78)	(0.80)
2	1.10	0.94	0.91	1.01	0.88	0.89	1.03	0.90	0.89
	(1.56)	(1.31)	(1.26)	(1.40)	(1.19)	(1.19)	(1.40)	(1.21)	(1.19)
3	1.20	0.99	0.99	1.16	1.10	1.03	1.16	1.01	1.07
	(1.70)	(1.39)	(1.36)	(1.60)	(1.51)	(1.41)	(1.63)	(1.38)	(1.45)
4	1.06	1.08	1.11	1.21	1.37	1.39	1.07	1.25	1.28
	(1.55)	(1.48)	(1.51)	(1.68)	(1.84)	(1.85)	(1.50)	(1.67)	(1.71)
5	2.24	2.00	1.85	3.46	2.53	2.23	2.98	2.39	2.16
	(2.82)	(2.60)	(2.43)	(3.71)	(3.12)	(2.85)	(3.47)	(3.06)	(2.80)
5-1	**1.36**	**1.10**	**0.95**	**2.88**	**1.93**	**1.61**	**2.46**	**1.78**	**1.53**
	(3.43)	**(4.07)**	**(4.19)**	**(4.71)**	**(4.73)**	**(5.06)**	**(5.32)**	**(5.42)**	**(5.32)**
SpearmanR	0.70	1.00	1.00	1.00	1.00	1.00	0.90	1.00	1.00
P-value	0.19	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00

This table reports the weekly performances of the LF, PE, and CM momentum under three identification windows including 30-day, 90-day, and 180-day, respectively. The LF link defines a directional relationship from the stock appearing in the news title (leaders) to the stock mentioned in the body text of the same news (followers) during the identification window. The CM link considers stocks appearing in the same sentence of a news article during the identification window as news co-mention connected stocks. The PE link defines two firms to be peer stocks with each other if they are only mentioned in the same news body during the identification window. The predictive signal of one focal stock is the average return of its connected stocks defined by each of the three links, weighted by the co-mention times (for LF momentum, the weight is the leading times). For each type of momentum, at the end of each trading day, all sample stocks are sorted quintiles based on the related predictive signal. Stocks are equal-weighted or value-weighted within each quintile portfolio. The long-short portfolio involves buying the highest group and selling the lowest group. All portfolios are held for one week and are rebalanced weekly. Panel A reports the mean returns of portfolios, while Panel B reports the Liu et al. (2019) Chinese four-factor adjusted alphas. The sample period is 2012 - 2020. All daily returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coefficient between the portfolio return and the serial number for each sorting. Newey and West (1987) adjusted t-statistics are shown in parentheses. Long-short returns/alphas with t-statistics higher than 2.00 are highlighted in bold.

Appendix

Appendix A. Descriptive Statistics for the News Data

Table 21: Descriptive statistics for the news data

Year	Mean	Std.	Max	Min
2006	97.23	59.60	252	1
2007	88.04	61.19	337	1
2008	70.72	28.88	152	3
2009	57.83	27.19	137	12
2010	61.06	33.66	135	1
2011	72.23	41.87	194	1
2012	84.29	52.16	183	1
2013	243.08	185.71	657	1
2014	336.04	212.59	699	1
2015	236.23	158.75	589	1
2016	197.29	131.52	493	1
2017	126.35	89.32	359	1
2018	156.37	105.31	467	1
2019	136.68	94.55	464	1
2020	334.87	313.69	1685	1
2021	521.84	392.11	2592	1

This table summarizes the daily number of news items for each year from 2006 to 2021.

Appendix B. Examples of the News and LF-only Links

In this section, we give two examples of news articles. Leader stocks are those appearing in the news titles and are colored in blue, while follower stocks are those only mentioned in the news bodies, and are colored in cyan.

Original news (in Chinese):

鲁抗医药 (600789)：A股增发精彩回放

鲁抗医药董事长章建辉先生致辞各位投资者朋友、各位网友：

大家好！欢迎大家参加山东鲁抗医药股份有限公司 A 股增发网上路演，在此，我代表公司董事会及公司全体员工对各位的关注与支持表示热烈欢迎和衷心感谢！山东鲁抗医药股份有限公司是国内抗生素生产基地之一，主营人用抗生素、半合成抗生素、动植物用抗生素、生物技术药品四大系列产品及抗生素相关制剂品种的生产、经营和销售。

...

问题：辛伐他汀确实是一种较有潜力的降血脂药物，但是上市公司中就有海正药业 (600267)、华东医药 (000963)生产，国内生产该类产品的厂家不在少数，公司如何看待这个问题?

林永彬：正如您所说，这个品种有许多厂家在生产，但我们更应该看到的是这类药物的巨大市场潜力。鲁抗医药对这个产品已经潜心研究了多年，而且本项目是从生物发酵生产洛伐他汀再合成该产品，我们掌握的技术是国内领先的，对该项目我们有充分的信心。

Translation version (in English):

Lukang Pharmaceutical (600789): a wonderful playback of the additional A-share issuance

Mr. Zhang Jianhui, Chairman of Lukang Pharmaceutical, delivers a speech:

Hello everyone! Welcome to participate in the online roadshow of Shandong Lukang Pharmaceutical Co., Ltd.. 's A-share additional issuance. Here, on behalf of the board of directors and all the staff of the company, I would like to express my warm welcome and heartfelt thanks to all of you for your attention and support! Shandong Lukang Pharmaceutical Co., Ltd. is one of the domestic antibiotic production bases, mainly engaged in the production, operation, and sales of four series of products and antibiotic-related preparations of antibiotics for human use, semi-synthetic antibiotics, antibiotics for animals and plants, biotechnology drugs.

...

Question: Simvastatin is indeed a kind of lipid-lowering drug with more potential, but among the listed companies, there are Haizheng Pharmaceutical (600267) and Huadong Pharmaceutical (000963), and there are not a few domestic manufacturers producing such products. How does the company view this problem?

Lin Yongbin: As you said, there are many manufacturers in the production of this variety, but what we should see more is the huge market potential of this kind of drug. Lukang Medicine has been studying this product for many years, and this project is to produce lovastatin by biological fermentation and resynthesis of this product. Our technology is leading in China, and we have full confidence in this project.

Original news (in Chinese):

澄星股份、华西村获省级高新技术企业认定

据江苏省科技厅 2002 年 12 月 27 日公告，澄星股份（600078）、华西村（000936）被认定为江苏省第 11 批高新技术企业。此外，宏图高科（600122）、国电南自（600268）、小天鹅（000418）、法尔胜（000890）和常林股份（600710）的参、控股子公司江苏宏图三胞科技发展有限公司、南京国电南自软件工程有限公司、无锡小天鹅精密铸造有限公司、江苏法尔胜光子有限公司、常州现代工程机械有限公司等也榜上有名。

Translation version (in English):

Chengxing Phosph-Chemicals、Huaxi Village were identified as provincial high-tech enterprises

According to the announcement of the Science and Technology Department of Jiangsu Province on December 27, 2002, Chengxing Phosph-Chemicals (600078) and Huaxi Village (000936) were identified as the 11th batch of high-tech enterprises in Jiangsu Province. In addition, the affiliated and controlling subsidiaries of Jiangsu Hongtu High Technology Co., Ltd. (600122), Guodian Nanjing Automation Co., Ltd. (600268), Little Swan Co., Ltd. (000418), Faersheng (Faston) Technology Co., Ltd. (000890), and Changlin Co., Ltd. (600710), including Jiangsu Macrotel Sanbao Technology Development Co., Ltd., Nanjing China National Electric South Software Engineering Co., Ltd., Wuxi Little Swan Precision Casting Co., Ltd., Jiangsu Farsheng Photonics Co., Ltd., and Changzhou Modern Engineering Machinery Co., Ltd., have also made it to the list.

Table 22: Examples of LF-only links

Date	Leader	Leader code	Follower	Follower code	Leading times	Events
10/1/2021	TCL	000100	Homa Appliance	002668	10	TCL acquired Homa Appliance and gained control.
3/1/2021	Ping An Group	601318	Ping An Bank	000001	4	Ping An Group is the largest controlling shareholder of Ping An Bank.
3/1/2021	Ping An Group	601318	China Fortune Land Development	600340	2	Ping An Group is the largest controlling shareholder of China Fortune Land Development.
9/1/2020	TCL	000100	Tianjin Printronics Circuit	002134	6	Tianjin Printronics Circuit is a wholly-owned subsidiary of TCL.
9/1/2020	TCL	000100	TCL Zhonghuan	002134	6	TCL Zhonghuan is a wholly-owned subsidiary of TCL.
11/1/2021	Ecovacs Robotics	603486	SIASUN Robot & Automation	300024	10	SIASUN Robot & Automation is the robot industry chain supplier of Ecovacs Robotics.
3/1/2020	ZTE	000063	SIASUN Robot & Automation	300024	2	SIASUN Robot & Automation is the robot industry chain supplier of ZTE.
6/1/2018	Midea Group	000333	SIASUN Robot & Automation	300024	5	SIASUN Robot & Automation is the robot industry chain supplier of Midea Group.
6/1/2020	Risen Energy	300118	CECEP Solar Energy	000591	2	CECEP Solar Energy is the solar cell module supplier of Risen Energy.
6/1/2020	LONGi	601012	CECEP Solar Energy	000591	4	CECEP Solar Energy is the solar cell module supplier of LONGi.

This table provides some examples of LF-only links, i.e., links that have never been identified as CM links during the identification window.

Appendix C. A Simple Illustration of the PE Link Construction

In this part, we give an example of constructing the **PE** matrix in a simplified situation. Suppose that there are only three stocks: stocks A, B, and C. At time t, during the past 90 days, A and B are co-mentioned for 3 times, A and C are co-mentioned for 2 times, and B and C are co-mentioned for 1 time. So we can get the news co-mention connection matrix:

$$\mathbf{CM}_t = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 3 & 2 \\ 3 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix} \end{array}.$$

During the same identification window, stock A has been led by C 2 times, and stock B has been led by A 1 time. So the LF matrix is [24]:

$$\mathbf{LF}_t = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array}.$$

Then,

$$\mathbf{LF^*}_t = \mathbf{LF}_t + \mathbf{LF}_t^{'} = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \end{array} + \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \end{array} = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 0 \\ 2 & 0 & 0 \end{pmatrix} \end{array}.$$

And

$$\mathbf{I}[\mathbf{LF^*}_t] = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array}.$$

Finally, we can get the PE matrix:

$$\mathbf{PE}_t = \mathbf{CM}_t \odot \mathbf{I}[\mathbf{LF^*}_t] = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{A} & \mathrm{B} & \mathrm{C} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}.$$

That is, only stock B and stock C have the PE link during this identification window.

[24] For the LF matrix, rows indicate followers, and columns indicate leaders.