Diamond Cuts Diamond: News Co-mention Momentum Spillover Prevails in China

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

September 30, 2023

Abstract

We conduct a comprehensive study on momentum spillovers in the Chinese stock market using various types of economic linkages. By developing a flexible and innovative algorithm to identify linkages among listed firms using millions of Chinese business news articles, we find that the news co-mention momentum spillover is significantly stronger compared to other forms of momentum spillovers. Using spanning tests and Fama-MacBeth regressions, we further show that the news co-mention momentum spillover unifies all different forms of momentum spillover effects in the Chinese stock market. Notably, the analyst co-coverage momentum spillover effect, which is the dominant species in the US stock market, is subsumed by the news co-mention momentum spillover effect in the Chinese stock market. We further explore the differences in the information content of links implied by news co-mentioning and other proxies. We suggest that the dominance of news co-mention momentum spillover over others can be attributed to two primary factors: comprehensive information and prompt updates.

Keywords: Economic Linkage, Big Textual Data, Momentum Spillover, News co-mention, Limited attention

JEL Classification: G11, G12, G14

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

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

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

1 Introduction

In the past decade, a new momentum-like anomaly known as \momentum spillovers" or \cross- rm momentum" has raised a lot of attention in the US stock market. A focal rm's stock price responds sluggishly to new information about economically linked rms, and as a result, the past returns of linked rms predict the future return of the focal rm. The literature documents the existence of such momentum spillover e ect via various types of economic linkages, including industry links (Moskowitz and Grinblatt, 1999; Hou, 2007), customer-supply links (Cohen and Frazzini, 2008), geographic links (Parsons et al., 2020; Jin and Li, 2020), technology links (Lee et al., 2019; Duan et al., 2022), news-implied links (Scherbina and Schlusche, 2013), concept links (Du et al. (2022)), and analyst co-coverage links (Ali and Hirshleifer, 2020), etc.

In the US market, the cross- rm momentum via analyst co-coverage uni es the \zoo" of other momentum spillover e ects (Ali and Hirshleifer, 2020). Unlike developed markets, the Chinese stock market is relatively young, and it possesses several distinctive features that make it particularly intriguing for research. The rst prominent feature of the Chinese stock market is the dominance of retail investors who are less sophisticated (Bailey et al., 2009). It's generally more costly for them to collect and process information, and they are more likely to be inattentive to new information about linked rms while making investing decisions. Consequently, we might expect a stronger lead-lag e ect among linked rms in the Chinese stock market. The second prominent feature of the A market is that investors trade in response to news quickly (Pan et al., 2016), which leads to its high turnover rate. Given these special features, two natural questions arise: whether various forms of momentum spillover e ects exist in the Chinese stock market and whether there exists a dominant species that uni es the \zoo" of all momentum spillover e ects.

In this paper, we conduct a comprehensive study on momentum spillovers in the Chinese stock market using various types of economic linkages. We nd that the momentum spillover e ect among rms linked through news co-mentioning (referred to as the news co-mention momentum spillover) is signi cantly stronger than others. Using spanning tests and Fama and MacBeth (1973) regressions, we show that all di erent forms of cross- rm momentum e ects can be uni ed by the news co-mention momentum spillover in the Chinese stock market. Notably, the analyst co-coverage momentum spillover e ect, which is the dominant species in the US stock market, is subsumed by the news co-mention momentum spillover e ect in the Chinese stock market. We further explore the di erences in the information content of links implied by news co-mentioning and other proxies. And we suggest that the dominance of news co-mention momentum spillover over others can be attributed to two primary factors: comprehensive information and prompt updates.

To be speci c, to construct the news co-mention linkages, we adopt two strategies. The rst one follows Scherbina and Schlusche (2013) and Ge et al. (2022), which de ne two rms as linked if they are co-mentioned in the same piece of news article during a pre-speci ed identi cation window. This identi cation strategy is referred to as*same_article*. One potential issue with this identi cation strategy is that two rms that happen to appear in the same article may appear in di erent sentences and are actually unrelated to each other. Given this concern, the second identi cation strategy we adopt is called*same_sentence*, which de nes two rms as linked if they are co-mentioned in the same sentence of the piece of news article during a pre-speci ed identi cation window.[1]

With the news-implied linkages, we proceed to test the existence of the news co-mention momentum spillover

[1] A similar link identi cation strategy is adopted in Schwenkler and Zheng (2019) and Schwenkler and Zheng (2021).

e ect. Duan et al. (2022) document that the momentum spillover e ects in the A market are more prominent for the weekly returns as the turnover rate is extremely high in the Chinese stock market. In addition, He et al. (2021) argue that Chinese investors may pay more attention to the past week's returns rather than the past month's returns. Given the above reasons, we mainly focus on momentum spillovers at the weekly frequency. At the end of each week, we construct a signal for each stock based on the weighted average returns of its linked stocks. We sort stocks into quintiles, and a long-short strategy could be constructed, which involves the purchase of stocks from the top quintile while shorting stocks from the bottom quintile. We construct eight such long-short trading strategies using combinations of the two identi cation strategies and four identi cation windows (window lengths including 3-, 6-, 9-, and 12-month),[2] and these trading strategies that are based on news co-mention momentum spillover all generate signi cant and positive abnormal returns. For instance, adopting same_sentence identi cation strategy and a 3-month identi cation window, the long-short return is 1.94% (t-statistic=5.33), and the corresponding Liu et al. (2019) four-factor (CH-4 for short later) adjusted alpha is 1.86% (t-statistic=5.23).[3] Moreover, we nd that the same_sentence strategy works better than the same_article strategy as the former yields larger abnormal returns with the same identi cation window. This nding aligns with our expectation that the same_sentence strategy is more e ective in identifying genuine links among rms. Another interesting nding is that the strength of news co-mention momentum spillover decreases steadily as the length of the identi cation window increases. This can be attributed to the fast updating of news, rendering stale information less useful in identifying relevant linkages.

We then compare the strength of news co-mention momentum spillover with various other momentum spillover e ects studied in the existing literature. The seven other cross- rm momentum e ects we examine include shared-analyst momentum, industry momentum, geographic momentum at the province and city levels, customer-supplier momentum, technology momentum, statistical momentum, and concept momentum. Similar to the previous case, for each linkage type, we construct signals for each stock based on the weighted average returns of its linked stocks and sort all stocks into quintiles. The long-short strategies exploiting most of those cross- rm momentum e ects also yield positive and statistically signi cant abnormal returns. However, these returns are noticeably smaller in both economic and statistical magnitudes compared to those obtained using news co-mentioning.

Given that the news co-mention momentum spillover e ect is signi cantly stronger than others, we follow Ali and Hirshleifer (2020) to use spanning tests and Fama and MacBeth (1973) regressions to investigate whether the news co-mention momentum spillover encompasses and uni es the various other momentum spillover e ects observed in the Chinese stock market.

To conduct the spanning tests, we construct the long-short factor returns based on news co-mentioning and the other six types of linkages discussed earlier.[4] Considering the superior performance of the same_sentence identi cation strategy, we focus on this strategy for the construction of the news co-mention momentum spillover factor. We construct four factors using identi cation windows of 3, 6, 9, and 12 months. For instance, we denote Sentence3 as the news co-mention momentum spillover factor based on identifying linkages using a 3-month window with the same_sentence strategy. Strikingly, the alphas of all other momentum spillover factors become

[2] The identi cation window cannot be too short as it would result in only a small number of rms with news-links.

[3] These returns are monthly returns converted from weekly ones.

[4] Due to the low quality of disclosed data regarding top customers in China, we do not consider the customer-supplier momentum spillover factor, which is often used as the representative cross- rm momentum in the US (e.g., Huang et al. (2021) and Huang et al. (2022)).

insigni cant, and some even turn negative when we add the news co-mention momentum spillover factor to the CH-4 factor model. For instance, the long-short strategy based on statistical momentum generates a CH-4 alpha of 1.34% (t-statistic=3.35), and its economic magnitude ranks second among all competitors. However, when we add the Sentence3 factor, its CH-4+Sentence3 model adjusted alpha decreases drastically to 0.17% (t-statistic=0.43). In contrast, none of the other six momentum factors can explain any of the four news co-mention momentum spillover factors (i.e., Sentence3, Sentence6, Sentence9, Sentence12). Additionally, among the four news co-mention momentum spillover factors, the factor constructed with a shorter identi cation window can explain the factors with longer identi cation windows, but not vice versa. The factor-spanning tests show that all di erent forms of cross- rm momentum e ects can be uni ed by the news co-mention momentum spillover in the Chinese stock market. The conclusion remains to be robust when we change the factor model, industry classi cation, and news co-mention type. Fama-MacBeth regressions further support this conclusion. Based on both union and intersection samples, after adding the news co-mention peer rm returns as an explanatory variable, the variables based on other cross- rm momentum e ects all become insigni cant.

We then explore the di erences in the information content of links implied by news co-mentioning and other proxies. Scherbina and Schlusche (2013), Schwenkler and Zheng (2019) nd that news co-mentioning reveals a wide range of economic linkages among companies, including same industry, business alliances, partnerships, banking and nancing, customer-supplier, production similarity, etc. To examine whether the news co-mention momentum spillover is stronger than others because news-implied linkages are more comprehensive, we conduct an exercise to analyze the within- and cross-industry momentum spillovers separately for each type of linkage. Firstly, we nd that, for the news co-mention linkage, the degree of overlap with industry information is relatively low compared to other types of linkages. On the other hand, the shared-analyst linkage shows the highest degree of overlap with the industry linkage. Next, for each linkage type, we decompose linked rms into industry peers and cross-industry peers, and we examine the two corresponding momentum spillover e ects separately. We nd that for news co-mentioning, the long-short portfolio based on within-industry and cross-industry momentum spillovers generate very similar returns and CH-4 adjusted alphas. In contrast, for other linkage types, the within-industry momentum spillover is noticeably stronger. For example, in the case of geographic and technology linkages, we nd that only the long-short portfolio based on the within-industry momentum spillover e ect generates a positive and statistically signi cant alpha. The cross-industry momentum spillover e ect does not yield signi cant abnormal returns. In the case of analyst co-coverage links, we nd that although its cross-industry momentum generates a positive alpha, it is marginally signi cant and only half the size of that obtained from the within-industry momentum. This exercise provides evidence that news co-mentioning reveals a wider range of economic links, and news-implied linkages are more comprehensive compared to other types of linkage proxies. Speci cally, news co-mentioning incorporates both industry information and valuable non-industry linkage information, which contributes to its strong predictive power for future returns.

In addition to containing comprehensive information, another advantage of news is its prompt update, which helps us to identify changes in linkages among rms in time. We keep track of the percentage change in di erent linkage networks over time, and we nd that the news-implied network updates faster than other networks. Every week, some links are added to the new network, and some stale links are removed. Due to the fast update, it is more challenging for investors to gather and process this information, which might explain why news co-mention momentum spillover is stronger. In contrast, other networks are more persistent, making it relatively easier for investors to take this information into account when making decisions.

After confirming that news co-mention momentum spillover unifies various momentum spillover effects and highlighting the advantages of news in identifying linkages, we further investigate the mechanism behind this news-based momentum spillover. We find that the news co-mention momentum is weaker among firms with higher analyst coverage, more analyst reports, larger float values, higher institutional holder proportions, less opacity, and less linkage complexity. These findings provide support for the limited attention explanation of news co-mention momentum[5]. Recent work by Huang et al. (2021) proposes a new behavioral explanation for the cross-firm momentum anomaly, suggesting that investors are slow to react to good news when the stock price is near its 52-week high and slow to react to bad news when the stock price is far from its 52-week high due to the anchoring effect. However, our empirical analysis does not support this explanation, possibly because the 52-week high effect itself is not empirically supported in the Chinese stock market (Zhang et al., 2019; Hou et al., 2023).

We conduct some additional heterogeneity analysis and robustness checks. In China, state-owned enterprises (SOEs) play a critical role in shaping the market dynamics and contributing to the overall economy. We conduct separate portfolio analyses for SOEs and non-SOEs, and interestingly, we find that the news co-mention momentum spillover is stronger for non-SOEs than SOEs. It is possible that the observed difference is due to media bias. Given that SOEs in China are politically related, it is plausible that the news co-mentioning about SOEs may be less precise in capturing peer firm relationships compared to non-SOEs. In the robustness part, we take transaction cost and shell effect into consideration. Following Fan et al. (2021), we consider a transaction cost of 16 bps for each trade (buying and selling combined). The long-short portfolio constructed based on news co-mention momentum generates a mean return of 1.29% (t-statistic=3.56), and the CH-4 adjusted alpha is 1.21% (t-statistic=3.42). In contrast, after taking transaction cost into consideration, trading strategies based on other cross-firm momentum effects all have insignificant alphas. This suggests that the news co-mention momentum strategy remains robust and profitable even when considering real-world transaction costs while others are not. To eliminate the potential shell effect in the Chinese stock market, we follow Liu et al. (2019) to drop firms with sizes at the bottom 30% of the market. Our key results remain unchanged.

This paper contributes to two strands of literature. Firstly, it provides new evidence on cross-firm momentum. Previous literature focus on testing the cross-firm momentum driven by different linkages in the US stock market (e.g., industry connection (Moskowitz and Grinblatt, 1999; Hou, 2007), supply chain (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Gemcay et al., 2015), conglomerate-standalone linkage (Cohen and Lou, 2012), alliance partnership (Cao et al., 2016), annual report similarity (Hoberg and Phillips, 2018), technology similarity (Lee et al., 2019), geographic proximity (Parsons et al., 2020; Jin and Li, 2020), analyst co-coverage (Ali and Hirshleifer, 2020), competition relationship (Eisdorfer et al., 2022), etc.). In this paper, we conduct a comprehensive study on momentum spillovers in the Chinese stock market using various types of economic linkages. It sheds light on the interplay between different linkages and their effects on stock returns, providing valuable insights for investors and researchers interested in cross-momentum strategies in emerging markets. Broadly speaking, this paper also contributes to the expanding literature on empirical asset pricing and anomalies in the Chinese stock market (see Jiang et al. (2018), Leippold et al. (2022), Jansen et al. (2021), Hou et al.

[5] Most previous studies attribute the momentum spillover effect to investors' limited attention to news from peer firms. See Burt and Hrdlicka (2021) for a detailed list of these studies. If this limited attention theory explains the news co-mention momentum in China, we would expect to see a weaker momentum spillover effect when information from peer firms is more transparent and easier to get.

(2023), and others).

The paper relates to the literature that extracts soft information from textual data. In recent years, there has been an explosion of empirical research in economics and nance that utilizes text as data. To mention a few examples, Ke et al. (2019) utilize information from news articles for predicting asset returns, Hu et al. (2021) extract information from Reddit to study the impact of social media on market dynamics, and Cong et al. (2019) develop a framework to generate 'textual factors' from large text datasets. News data has received signi cant attention among various alternative data sources, and there is a growing interest in using news coverage to identify linkages among rms. Studies such as Scherbina and Schlusche (2013) have demonstrated the usefulness of rm linkages identi ed 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 rms, and they also applied the same identi cation strategy to establish crypto peers (Schwenkler and Zheng (2021)). Guo et al. (2017) have explored the association between news co-mentioning and investor attention spillovers. This paper adopts two di erent identi cation strategies to infer rm linkages from news, and we show that sentence co-mentioning strategy is more e ective in identifying genuine links among rms. Unlike the ndings of Scherbina and Schlusche (2013), who use article co-mentioning to identify linkages and found that trading strategies based on the predictability of linked rms' past returns may not be pro table when considering transaction costs, we demonstrate that our strategy, based on sentence co-mentioning to infer links, remains robust and pro table even when accounting for real-world transaction costs. In addition, we show that due to the unique advantages of news information, all di erent forms of cross- rm momentum e ects can be uni ed by the news co-mention momentum spillover in the Chinese stock market.

The rest of the paper is organized as follows. In Section 2, we introduce the data and the construction of di erent linkages. Section 3 conducts the portfolio analysis for each type of momentum spillover e ect. In Section 4, we examine the unifying e ect of the news co-mention momentum spillover and explore the di erences in the information content of links implied by news co-mentioning and other proxies. Section 5 investigates the mechanism underlying the news co-mention momentum. In Section 6, we conduct some further analysis and robustness checks. Section 7 concludes. Additional materials are given in Appendix.

2 Data and summary statistics

The sample stocks used in this paper include 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 (ST) shares are excluded[6]. In the robustness analysis, we also exclude rms that have the bottom 30% capitalization from the sample. This is done to prevent potential biases related to the shell e ect in the Chinese stock market, as highlighted in the study by Liu et al. (2019). The full sample period for our analysis spans from 2006 to 2020. Due to data availability issues speci c to each linkage type, the sample periods may vary and be shorter when conducting analysis for di erent types of linkages. Stock trading data, nancial statements, and risk-free interest rates (one-year deposit rate) are from CSMAR. The CH-4 factor data is obtained from Robert

[6] Since April 22, 1998, the Shanghai and Shenzhen stock exchanges have implemented Special Treatment (ST) for the stock trading of listed rms with abnormal nancial conditions. ST shares are pre xed with "ST" and are considered to have extremely high risk. Therefore, they are often excluded from research about the Chinese stock market.

F. Stambaugh's website.[7]

In later subsections, we provide details on the data sources and construction methodology for each type of economic linkage considered in the study, as well as the corresponding predictive signals derived from these linkages. We first discuss the news co-mention links, which are our main focus. Then, we provide brief descriptions of seven other proxies for inter-firm linkages commonly studied in the literature, including the analyst co-coverage (Ali and Hirshleifer, 2020), customer-supplier relationship (Cohen and Frazzini, 2008), industry classification(Hou, 2007), geographic proximity (Parsons et al., 2020), technology affinity (Lee et al., 2019), statistical similarity of firm features (He et al., 2021), and concept category (Du et al., 2022).

2.1 News Co-mention Linkages

We use millions of news articles from the Financial Text Intelligent Analysis Platform of RESSET and the Juyuan Database spanning from 2006 to 2020.[8] We filtered out 1,138,247 news articles that mentioned at least one listed firm in the A market. Table 14 from Appendix A presents a summary of the daily basic information of the news data since 2006. Prior to 2012, the news data is 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 2020 when the news data quality is high. For robustness check, we also consider different sample periods.

We adopt two strategies to construct the news co-mention linkages. The first identification strategy is referred to as*same_article*, which defines two firms as linked if they are co-mentioned in the same piece of news article during a pre-specified identification window. One potential issue with this identification strategy is that two firms that happen to appear in the same article may appear in different sentences and are actually unrelated to each other. Given this concern, the second identification strategy we adopt is called*same_sentence*, which defines two firms as linked if they are co-mentioned in the same sentence of the same piece of news article during a pre-specified identification window.

For both identification strategies, we consider identification windows of various lengths, including 3 months, 6 months, 9 months, and 12 months. The identification window cannot be too short or too long in order to strike a balance between capturing an adequate amount of news co-mentions and the inclusion of very stale information. Short windows (e.g., 1-month) result in a small number of stocks with news links, while long windows (e.g., over one year) include outdated news with limited predictability. At the end of each week, for a focal firm i, we calculate the news-based predictive signal as the average excess return of linked firms weighted by the number of co-mentions during the identification window. To be precise, we have eight news-based signals with different identification strategies and window lengths. For example, the signal constructed based on*same_sentence*and 3-month window is denoted as*Sentence3_Rtn*, and the corresponding trading strategy is referred to as*Sentence3*.

[7] The website provides monthly and daily CH-4 factor data. We construct weekly factors by using cumulative daily factor returns in one week (i.e., at the end of each trading week, the weekly CH-4 factor is the cumulative return of daily CH-4 factor returns from the first trading day to the last trading day of the week).

[8] We access the data from the Financial Engineering Laboratory of Peking University.

Table 1: Summary statistics

Link type	Variables	Count	Mean	Std.	Min.	Median	Max.
Sentence_3	# Stocks	464	1100	280	715	996	1840
	# Peer firms	510429	9	16	1	3	235
	Value	510429	22.45	71.08	0.27	7.19	2509.88
	Peer firm return	510429	0.0050	0.0697	-0.4101	0.0034	24.1232
Sentence_6	# Stocks	464	1515	231	1164	1462	2324
	# Peer firms	702802	12	21	1	4	311
	Value	702802	19.12	61.36	0.27	6.76	2509.88
	Peer firm return	702802	0.0044	0.0641	-0.4102	0.0036	24.1232
Sentence_9	# Stocks	464	1758	187	1316	1756	2502
	# Peer firms	815530	15	25	1	5	321
	Value	815530	17.73	57.27	0.27	6.56	2509.88
	Peer firm return	815530	0.0038	0.0555	-0.4101	0.0036	1.5622
Sentence_12	# Stocks	464	1925	199	1437	1915	2613
	# Peer firms	893012	17	28	1	6	355
	Value	893012	16.91	54.89	0.27	6.41	2509.88
	Peer firm return	893012	0.0034	0.0540	-0.4103	0.0035	1.0741
Article_3	# Stocks	464	1332	293	903	1234	2223
	# Peer firms	618272	16	32	1	5	454
	Value	618272	20.41	65.11	0.27	6.96	2509.88
	Peer firm return	618272	0.0047	0.0559	-0.4101	0.0043	3.3824
Article_6	# Stocks	464	1750	233	1355	1742	2704
	# Peer firms	812053	23	42	1	8	631
	Value	812053	17.79	57.39	0.27	6.58	2509.88
	Peer firm return	812053	0.0041	0.0526	-0.4101	0.0044	0.8516
Article_9	# Stocks	464	1976	229	1478	1952	2816
	# Peer firms	917064	29	51	1	10	757
	Value	917064	16.69	54.22	0.27	6.38	2509.88
	Peer firm return	917064	0.0036	0.0507	-0.4097	0.0042	1.0741
Article_12	# Stocks	464	2122	278	1569	2121	2891
	# Peer firms	984555	35	59	1	12	867
	Value	984555	16.06	52.43	0.27	6.24	2509.88
	Peer firm return	984555	0.0033	0.0493	-0.4103	0.0041	1.0741
Analyst	# Stocks	768	1326	348	476	1429	1872
	# Peer firms	1018102	98	84	1	75	609
	Value	1018102	15.45	57.09	0.18	6.60	6285.82
	Peer firm return	1018102	0.0032	0.0445	-0.4099	0.0049	0.9588
Customer	# Stocks	566	109	56	24	71	197
	# Peer firms	61428	1	0	1	1	5
	Value	61428	8.37	20.68	0.08	4.53	517.98
	Peer firm return	61428	0.0025	0.0580	-0.3662	0.0008	2.6103
Industry	# Stocks	768	2336	795	1048	2313	3893
	# Peer firms	1794095	130	83	2	110	364
	Value	1794095	13.85	59.28	0.11	4.88	6285.82
	Peer firm return	1794095	0.0039	0.0463	-0.2713	0.0058	0.5055
Geographic	# Stocks	768	2355	783	1088	2329	3892
	# Peer firms	1808617	201	162	3	157	651
	Value	1808617	13.84	59.05	0.09	4.88	6285.82
	Peer firm return	1808617	0.0040	0.0445	-0.2885	0.0063	1.0308
Technology	# Stocks	768	1020	513	318	954	2117
	# Peer firms	783093	680	501	1	599	2109
	Value	783093	17.99	82.98	0.08	5.27	6285.82
	Peer firm return	783093	0.0040	0.0437	-0.3096	0.0064	0.9269
Statistical	# Stocks	768	2015	704	988	2014	3548
	# Peer firms	1547320	263	79	115	240	390
	Value	1547320	14.19	56.96	0.09	5.03	2619.90
	Peer firm return	1547320	0.0032	0.0453	-0.2879	0.0052	0.2410
Concept	# Stocks	227	2970	497	2044	3028	3821
	# Peer firms	674244	433	504	1	260	2704
	Value	674244	17.14	60.78	0.48	6.21	2509.88
	Peer firm return	674244	0.0005	0.0340	-0.1711	0.0006	0.6113

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 news co-mention linkages are based on either same_sentence or same_article strategy, using identification windows of 3 months, 6 months, 9 months, and 12 months. The industry linkage is based on the Shenwan-1 classification. The geographic linkage is at the province level. The sample period of news co-mention momentum is 2012-2020. The sample period of customer-supplier momentum is 2010-2020. The sample period of concept momentum is Aug. 2016-2020. The sample periods for other momentums are 2006-2020.

2.2 Other Linkages

Next, we will describe the construction of other types of linkages in order to conduct a comprehensive study on momentum spillovers in the Chinese stock market.

2.2.1 Analyst Co-coverage Linkages

The analyst reports data used in our study are from the Chinese Research Data Services Platform (CNRDS). This platform functions similarly to Wharton Research Data Services (WRDS) and provides access to a wide range of Chinese research data in the elds of nance and economics. The sample period for our study is from January 2005 to December 2020, which covers the earliest available analyst prediction data on the CNRDS platform. There is a total of 1,478,413 pieces of analyst predictions over the sample period. After removing data with null values and eliminating duplicates based on analyst codes, report dates, and stock codes, we are left with 530,696 unique pieces of analyst prediction data for the identi cation of shared-analyst linkage.

We adopt the approach outlined by Ali and Hirshleifer (2020) to identify shared-analyst linkages and compute the shared-analyst peer rm return, denoted as Analyst_Rtn. At the end of each trading week, two rms are de ned as the shared-analyst peer rms if they are co-covered by the same analyst teams in the past 12 months.[9] At the end of each week, for a focal rm i, the shared-analyst peer rm return, Analyst_Rtn_{it}, is the average excess return of analyst peer rms, weighted by the number of shared analyst teams. Since a one-year identi cation window is required to identify the shared-analyst linkages, the predictive signals can be computed from the beginning of 2006.

2.2.2 Industry Linkages

There are three frequently used industry classi cation systems in China, namely CSRC, CITIC, and Shenwan.[10] The historical information about the industry classi cations used in the study is obtained from the RESSET database, where we collect yearly data for each classi cation level of CSRC, CITIC, and Shenwan. We mainly focus on the primary classi cation of Shenwan as it is the most popular classi cation system in the Chinese nancial sector. At the end of each week, for a focal rm i, the industry peer rm return, denoted as Industry_Rtn_{it}, is calculated as the equal-weighted average excess return of all other stocks from the same industry.

2.2.3 Customer-supplier Linakges

We follow Cohen and Frazzini (2008) and extract the information about the top-5 customers for all listed rms from their annual reports. However, for Chinese listed rms, the disclosure of the real names of their top customers is not mandated by China Securities Regulatory Commission (CSRC). Due to the voluntary nature of the disclosure, many rms choose to use digits, English letters, or Chinese numbers to represent their customers

[9] In the Chinese stock market, analyst reports are often produced by a team of analysts rather than individual analysts. In our study, we consider all analysts within a team as a cohesive unit and do not distinguish them as individual analysts. In other words, two stocks are de ned as shared-analyst peer rms only if they are covered by the same analyst team. This approach acknowledges that analyst reports are collaborative e orts that incorporate the ideas and insights of the entire team.

[10] In 2020, for the CSRC system, there are 19 primary industries and 81 secondary industries in the sample. For the CITIC system, there are 29 primary industries, 83 secondary industries, and 188 tertiary industries in the sample. For the Shenwan system, there are 28 primary industries, 104 secondary industries, and 227 tertiary industries in the sample.

instead of providing their real names or stock codes. This makes it challenging to identify rms' customers using the dataset. Moreover, some disclosed customers are not listed rms (they can be local governments, schools, and individuals). As a result, the number of customer-supplier linkages identi ed is much smaller than other linkages. Summary statistics from Table 1 show that for each trading week, on average, there are only 109 sample stocks whose customers can be clearly identi ed. To ensure an adequate number of sample rms have customer-supplier linkages, we limit the sample period from January 2010 to December 2020. To avoid any potential look-ahead bias, we set a lag of six months between the scal ending date and the portfolio formation date.[11] At the end of each week, for a focal rm i, the customer peer rm return Customer_Rtn_{it} is computed as the weighted average value of its customers' excess returns.

2.2.4 Geographic Linkages

According to the ndings of Parsons et al. (2020), the stock returns of a rm can be predicted by the returns of its peer rms that are headquartered in the same geographic location. In our study, we de ne geographic peer rms at both the province level and the city level in China. The historical data about headquarters locations are obtained from CSMAR.[12] There are 32 provincial districts (including four municipalities, ve provincial autonomous regions, and 23 provinces) and 399 cities (including four municipalities). In our analysis, we consider all rms located within the same province or city as geographic peers (i.e., two types of geographic linkages). At the end of each week, for a focal rm i, the geographic peer rm return $\text{Geographic\_Rtn}_{it}$ is calculated as the equal-weighted excess return of all of its geographic neighbor rms.

2.2.5 Technology Linkages

We follow the methodology in Lee et al. (2019) and de ne the technology peer rms based on technology similarity. The patent data is from the CCER database which provides detailed patent information for listed rms, including the International Patent Classi cation (IPC) patent class and the patent grant date. Technology similarity between rm i and j at time t is computed as

$$\text{TECH}_{ijt} = \frac{T_{it}\, T_{jt}^{0}}{(T_{it}\, T_{it}^{0})^{1=2}\ T_{jt}\, T_{jt}^{0}\ ^{1=2}};$$

$T_{it} = (T_{it\,1}; T_{it\,2};\ \ ; T_{itm})$ is the proportion vector of patents among each class in IPC in the ve years preceding time t for rm i, where m is the number of all classi cations. At the end of each week, for a focal rm i, the technology peer rm return, $\text{Technology\_Rtn}_{it}$, is the average return of its technology peer rms weighted by their technology similarity. To avoid the look-ahead bias, we set a lag of at least six months between the scal-year end and the portfolio formation date.

2.2.6 Statistical Linkages

He et al. (2021) nd that a rm's return can be predicted by the past return of rms that are similar to it on some particular characteristics. In line with their work, we de ne statistical peer rms as stocks that have the

[11] For example, for July to December 2020, we use the customer information disclosed in the annual report of 2019, while for January to June 2020, we use the customer information disclosed in the annual report of 2018.

[12] In the context of China, rms are typically associated with two geographical locations: the location of their headquarters and the place of registration. In most cases, these two locations are the same. However, if there is a di erence between the two, we prioritize the location of the headquarters. If the information regarding the headquarters location is not available but the registration address is, we use the registration address as a substitute.

closest Euclidean distance on five firm-level characteristics: closed stock price (P), firm size ($SIZE$), book-to-market ratio (BM), ROE (ROE), and assets growth (AG).[13] The distance between firm i and firm j is be computed as:

$$d_{ij} = \sqrt{(P_i - P_j)^2 + (SIZE_i - SIZE_j)^2 + (BM_i - BM_j)^2 + (ROE_i - ROE_j^2) + (AG_i - AG_j)^2}.$$

At the end of each week, for a focal firm i, we calculate the distances of all other firms with it, and the top 10% of firms with the smallest distances are considered as the statistical peers of firm i.[14] The statistical peer firm return $Statistical\_Rtn_{it}$ is then calculated as the equal-weighted excess return of its statistical peer firms.

2.2.7 Concept Linakges

In the stock market, one "concept" refers to a group of stocks that share a specific trend or topic, such as new energy or e-commerce. Following Du et al. (2022), two firms are considered concept peers if they belong to the same concept. The historical records of concepts and constituent stocks used in our study are obtained from the publicly accessible RESSET database.[15] Due to data availability, the sample period for concept momentum analysis in our study spans from August 2016 to December 2020. During this period, we identify a total of 336 concepts. Table 1 shows that for each week during the sample period, there are 2,970 sample stocks with concept linkages on average. At the end of each week, for a focal firm i, the concept peer firm return $Concept\_Rtn_{it}$ is the weighted average returns of its concept peers, where the weight is the number of common concepts.

2.3 Summary Statistics

Table 1 shows the summary statistics for different linkage types. Under $Sentence\_3$, each focal firm has nine peer firms on average, fewer than 16 peers from $Article\_3$. This aligns with our expectations as the $same\_sentence$ strategy removes the potential noise links from the $same\_article$ strategy, resulting in fewer links identified. Furthermore, the number of peer firms identified increases with the length of the identification window. For other linkage types (except customer-supplier), we generally observe a higher number of links, and each sample stock tends to have more peers on average.

[13] To avoid the look-ahead bias, we use accounting variables (BM, ROE, and AG) that are lagged by at least one year. For example, from January to June, the variables are based on the fiscal report two years ago, while from July to December, they are based on the fiscal report one year ago. For $SIZE$, the total market value at the end of June last year is used for the first half of the year, while the total market value at the end of June this year is used for the second half of the year. P is the weekly closed price at the end of the trading week. Each of the five variables is standardized by its cross-sectional mean and standard deviation.

[14] Different from He et al. (2021), where 50 smallest distances are used to define similar stocks in the US market, it may be more appropriate to use a percentage rather than a fixed number to define similar stocks in the Chinese market. This is because the number of A-shares in China has significantly increased over the past decades. By using a percentage of the sample size, we can maintain a consistent criterion for defining similar stocks regardless of changes in the overall market size.

[15] We acknowledge that there may be differences in the data sources and quality between our study and Du et al. (2022). They spent a lot of effort scraping the concept data from the Joinquant database. Since concept momentum is not our main focus, we rely on publicly available data, which may have limitations in terms of coverage and quality.

3 Portfolio Analysis

In this section, we examine the relationship between past peer firm return signals computed using various types of linkages and future stock returns. At the end of each trading week, all sample stocks are sorted into quintiles according to peer firm return signals computed using various linkages. Stocks are equal-weighted within each quintile group. 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. In addition to reporting the average return over the risk-free rate, we also calculate the alpha using the CH-4-factor model. The weekly returns and alphas are then converted to a monthly frequency for better comparability with existing literature.

3.1 News Co-mention Momentum Spillover

We first report the main results for news co-mention momentum spillovers. Table 2 presents the results of eight trading strategies based on different identification strategies (same_sentence and same_article) and window lengths (3 months, 6 months, 9 months, and 12 months). Panel A and Panel B show the excess returns, and CH-4 adjusted alphas of each portfolio, respectively.

Overall, the eight long-short portfolios that exploit news co-mention momentum spillovers exhibit statistically and economically significant positive excess returns, which can not be explained by the CH-4 factors, including the market, size, value, and abnormal turnover rate. Among the eight strategies, Sentence_3 performs the best and yields a monthly return of 1.94% (t-statistic=5.33), and even the worst performing one (Article_12) yields a monthly return of 1.36% (t-statistic=4.49). The Spearman correlation coefficients between the rank of news-based peer firm returns and the long-only portfolio returns are equal to one or close to one. This indicates a positive and monotonic relationship between the two variables, suggesting that higher news-based peer firm returns are associated with higher long-only portfolio returns.

When comparing the eight news-based trading strategies that employ different identification strategies and identification window lengths, we observe that while the overall news co-mention momentum spillover effect is strong, there are substantial differences among the eight combinations. First, give the same identification window length, same_sentence strategy always performs better than same_article strategy. Taking the 3-month windows as an example, Sentence_3 strategy generates a long-short return and alpha of 1.94% (t-statistics=5.33) and 1.86% (t-statistic=5.23) respectively, higher than 1.63% (t-statistic=1.63) and 1.58% (t-statistic=5.16) for that of Article_3. The results are similar for other identification windows, indicating that same_sentence might be more effective in identifying genuine links among firms. This is because when two firms appear in the same article but in different sentences, there are chances that they are unrelated to each other.

Table 2: The news co-mention momentum spillover: portfolio sorting results

Panel A: Excess returns								
	same_sentence				same_article			
Identification windows	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month
1 (Low)	0.74	0.81	0.85	0.84	0.89	0.88	0.85	0.84
	(0.88)	(0.96)	(1.02)	(1.00)	(1.07)	(1.06)	(1.02)	(1.01)
2	1.06	1.06	1.01	0.94	1.08	1.06	1.01	1.00
	(1.32)	(1.31)	(1.23)	(1.15)	(1.32)	(1.29)	(1.23)	(1.21)
3	1.11	1.16	1.09	1.15	1.20	1.15	1.12	1.09
	(1.38)	(1.42)	(1.34)	(1.38)	(1.49)	(1.40)	(1.35)	(1.31)
4	1.38	1.44	1.38	1.37	1.33	1.40	1.39	1.41
	(1.65)	(1.71)	(1.65)	(1.63)	(1.60)	(1.67)	(1.68)	(1.68)
5 (High)	2.68	2.38	2.30	2.22	2.54	2.34	2.26	2.21
	(2.96)	(2.66)	(2.57)	(2.48)	(2.84)	(2.66)	(2.52)	(2.49)
5-1	1.94	1.55	1.44	1.37	1.63	1.45	1.40	1.36
	(5.33)	(4.97)	(4.68)	(4.49)	(5.30)	(5.14)	(4.62)	(4.49)
SpearmanR	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
Panel B: CH-4 adjusted alphas								
	same_sentence				same_article			
Identification windows	3-month	6-month	9-month	12-month	3-month	6-month	9-month	12-month
1 (Low)	0.41	0.48	0.51	0.49	0.55	0.53	0.51	0.50
	(0.49)	(0.56)	(0.61)	(0.59)	(0.66)	(0.63)	(0.61)	(0.60)
2	0.78	0.74	0.69	0.61	0.77	0.74	0.67	0.66
	(0.98)	(0.92)	(0.84)	(0.75)	(0.95)	(0.91)	(0.82)	(0.80)
3	0.77	0.81	0.71	0.76	0.89	0.81	0.77	0.73
	(0.98)	(1.01)	(0.88)	(0.94)	(1.11)	(1.00)	(0.94)	(0.88)
4	1.06	1.10	1.06	1.04	0.99	1.06	1.04	1.07
	(1.29)	(1.35)	(1.30)	(1.27)	(1.21)	(1.30)	(1.28)	(1.29)
5 (High)	2.27	2.00	1.92	1.84	2.14	1.95	1.87	1.83
	(2.66)	(2.38)	(2.26)	(2.17)	(2.52)	(2.34)	(2.21)	(2.16)
5-1	1.86	1.51	1.41	1.35	1.58	1.42	1.35	1.32
	(5.23)	(4.89)	(4.59)	(4.38)	(5.16)	(4.98)	(4.45)	(4.35)
SpearmanR	0.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00
P-value	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00

This table reports the portfolio sorting results of the news co-mention momentum spillover effect of the same_sentence type and same_article type under the four identification windows, respectively. Same_sentence strategy considers firms appearing in the same sentence of the same news article during a given identification window as news co-mention peers, while the same_article strategy considers firms appearing in the same article during a given identification window as news co-mention peers. The identification windows include 3-month, 6-month, 9-month, and 12-month. The news co-mention peer firm return of a focal firm is computed as the average excess return of news co-mention peer firms of the focal firm during the identification window, weighted by the number of co-mentions. At the end of each trading week, all sample stocks are sorted quintiles based on the news co-mention peer firm returns. Within each quintile group, the stocks are equally weighted. 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. The sample period is 2012-2020. Panel A reports the excess returns of portfolios, and Panel B reports the intercepts (alphas) of the regression of the returns on CH-4 factors (market, size, value, abnormal turnover rate) (Liu et al., 2019). All weekly 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.

In addition, given the link identification strategy, the returns of the long-short portfolio decrease as the length of the identification window increases. Unlike other linkages such as industry links and technology links, which tend to be more stable and persistent in the long term, news is highly time-sensitive and reflects recent information from the market. News-based linkages capture the immediate market sentiment, investor behaviors, and breaking news events, which can quickly change over time. As a result, the news linkages are less persistent compared to other linkages and can exhibit rapid shifts as new information becomes available. The use of a long identification window may include outdated news that is no longer relevant. This might explain the observed results. However, a shorter identification window results in a smaller number of sample stocks with news-implied linkages. For instance, if we adopt *same_sentence* strategy and use a 1-month window to identify the links, there are only 500 sample stocks on average each week that have news-implied links. To strike a balance between the performance and the sample size, the shortest identification window is 3 months. Since Sentence3 strategy generates the largest returns among the eight candidates, for later comparison with other momentum spillover effects, we use the Sentence3 strategy.

3.2 Other Momentum Spillovers

Table 3 reports the portfolio sorting results for other types of linkages, including the analyst co-coverage, industry, geographic proximity, customer-supplier, technology, statistical similarity, and concept. For better comparison, the result of Sentence3 is included as the first column. Panel A shows the mean excess return, and the CH-4 adjusted alpha of each portfolio is reported in Panel B.

Most of these cross-momentum anomalies observed in the US market also exist in the Chinese stock market. In particular, the momentum spillover via statistical linkages yields a monthly long-short average excess return and alpha of 1.36% (t-statistic=3.93) and 1.34% (t-statistic=3.35), respectively, which is the second-highest among all strategies, following the Sentence3 strategy. The analyst co-coverage momentum spillover was found to unify all momentum spillover effects in the US market. In the Chinese stock market, a long-short portfolio based on that yields an average return of 0.86% (t-statistic=3.33) and CH-4 alpha of 0.85% (t-statistic=2.97). Long-short average return and alpha based on industry momentum using Shenwan-1 industry classification are 0.77% (t-statistic=3.25) and CH-4 alpha of 0.76% (t-statistic=3.00), respectively. The technology momentum generates a slightly lower average return and alpha of 0.58% (t-statistic=2.73) and 0.60% (t-statistic=2.64). Although we have a short sample for concept momentum, it produces a mean return of 0.81% (t-statistic=2.25) per month.[16]

Among all the cross-momentum strategies examined, only the customer momentum strategy does not yield positive returns or CH-4 alpha. We attribute the poor performance of the customer momentum strategy to the low quality and limited availability of disclosed customer data for listed firms in China. As previously mentioned, only a small number of firms choose to disclose the real names of their customers, and the majority of disclosed customers are not listed firms. Consequently, the sample that we can use to do portfolio analysis is small. To avoid potential bias resulting from this small sample size, we exclude the customer momentum in later analysis.

[16] Unlike the results reported by Du et al. (2022), our concept momentum strategy generates insignificant alpha after adjusting for the CH-4 factor. We believe that the discrepancy in results can be attributed to the difference in data sources related to concepts. They have utilized web-crawling technology to obtain additional and longer historical records about concepts, which may have contributed to their significant findings.

Table 3: Other momentum spillovers: portfolio sorting results

Panel A: Excess returns									
	News co-mention	Analyst	Industry	Province	City	Customer	Technology	Statistical	Concept
1 (Low)	0.74	1.18	1.35	1.51	1.61	0.92	1.54	0.91	-0.45
	(0.88)	(1.59)	(1.82)	(2.06)	(2.15)	(1.19)	(2.05)	(1.22)	(-0.50)
2	1.06	1.28	1.52	1.64	1.67	0.22	1.66	1.35	-0.15
	(1.32)	(1.80)	(2.05)	(2.21)	(2.25)	(0.27)	(2.21)	(1.75)	(-0.17)
3	1.11	1.61	1.74	1.77	1.78	0.17	1.94	1.71	0.10
	(1.38)	(2.24)	(2.28)	(2.37)	(2.41)	(0.22)	(2.56)	(2.25)	(0.11)
4	1.38	2.03	1.90	1.84	1.81	0.95	1.92	2.09	0.21
	(1.65)	(2.81)	(2.49)	(2.44)	(2.40)	(1.20)	(2.52)	(2.79)	(0.25)
5 (High)	2.68	2.04	2.12	1.86	1.79	0.38	2.12	2.28	0.36
	(2.96)	(2.74)	(2.74)	(2.42)	(2.35)	(0.48)	(2.72)	(3.03)	(0.44)
5-1	1.94	0.86	0.77	0.35	0.18	-0.53	0.58	1.36	0.81
	(5.33)	(3.33)	(3.25)	(2.77)	(2.05)	(-1.46)	(2.73)	(3.93)	(2.25)
SpearmanR	1.00	1.00	1.00	1.00	0.90	0.10	0.90	1.00	1.00
P value	0.00	0.00	0.00	0.00	0.04	0.87	0.04	0.00	0.00
Panel B: CH-4 adjusted alphas									
	News co-mention	Analyst	Industry	Province	City	Customer	Technology	Statistical	Concept
1 (Low)	0.41	0.82	0.98	1.13	1.23	0.72	1.11	0.58	-0.40
	(0.49)	(1.09)	(1.29)	(1.52)	(1.61)	(0.94)	(1.45)	(0.75)	(-0.44)
2	0.78	0.92	1.12	1.24	1.31	0.16	1.26	0.95	-0.11
	(0.98)	(1.26)	(1.48)	(1.64)	(1.74)	(0.21)	(1.65)	(1.22)	(-0.13)
3	0.77	1.29	1.32	1.39	1.37	0.15	1.61	1.29	0.10
	(0.98)	(1.75)	(1.70)	(1.82)	(1.80)	(0.19)	(2.10)	(1.67)	(0.11)
4	1.06	1.72	1.48	1.42	1.43	0.67	1.54	1.65	0.19
	(1.29)	(2.38)	(1.94)	(1.84)	(1.87)	(0.87)	(1.97)	(2.17)	(0.21)
5 (High)	2.27	1.68	1.74	1.48	1.36	0.12	1.71	1.93	0.33
	(2.66)	(2.21)	(2.20)	(1.89)	(1.75)	(0.16)	(2.14)	(2.51)	(0.39)
5-1	1.86	0.85	0.76	0.34	0.13	-0.60	0.60	1.34	0.74
	(5.23)	(2.97)	(3.00)	(2.56)	(1.37)	(-1.43)	(2.64)	(3.35)	(1.89)
SpearmanR	0.90	0.90	1.00	1.00	0.70	-0.70	0.90	1.00	1.00
P value	0.04	0.04	0.00	0.00	0.19	0.19	0.04	0.00	0.00

This table reports the portfolio sorting performance of the shared-analyst momentum, industry momentum, geographic momentum, customer momentum, technology momentum, statistical momentum, and concept momentum. For comparison, the result for the news co-mention momentum of same_sentence type under the 3-month identi cation window (Sentence_3) is shown in the rst column. The industry momentum spillover is based on the rst level of Shenwan classi cation (Shenwan-1), and the results are shown in column 2. The geographic momentum spillover is constructed at both the province and city levels, and the results are shown in columns 3 and 4, respectively. Columns 5-8 present the results for customer momentum, technology momentum, statistical momentum, and concept momentum, respectively. Due to data quality, the sample period of news co-mention momentum is 2012-2020, the sample period of customer momentum is 2010-2020, the sample period of concept momentum is Aug. 2016-2020, and the sample periods for other momentums are 2006-2020. For each linkage type, at the end of each trading week, we rst calculate the peer returns of all focal rms and then sort all stocks into quintiles based on the peer rm return. Within each quintile group, the stocks are equally weighted. 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 gives the excess returns of portfolios, and Panel B presents the intercepts of the regression of the returns on CH-4 factors (Liu et al., 2019) (market, size, value, abnormal turnover rate). All weekly returns and alphas are converted into monthly percentages using compound interest. SpearmanR reports the Spearman correlation coe cient 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.

In the Chinese stock market, the geographic momentum is not strong. Province-level momentum is statistically signi cant but the long-short strategy based on that yields only a small average return of 0.35% per month. City-level momentum is even weaker and is not signi cance after controlling for the CH-4 factors.

Overall, the news co-mention momentum spillover e ect is stronger than other types of momentum spillovers in the Chinese stock market. A natural question then arises: whether the news co-mention momentum spillover encompasses and uni es the various other momentum spillover e ects.

4 The Unifying E ect

In this section, we examine whether the news co-mention momentum spillover e ect encompasses and uni es the various other momentum spillover e ects observed in the Chinese stock market. To do so, we follow Ali and Hirshleifer (2020) to use spanning tests and Fama and MacBeth (1973) regressions. In addition, we investigate the information content di erences between news co-mentioning and other proxies to gain further insights into the features of news-implied linkages.

4.1 Factor Spanning Tests

To test whether the anomaly returns of di erent types of momentum spillovers can be explained by each other, we construct an augmented CH-4 model as follows:

$$R_t = \quad + \quad _{MKT}\, MKT_t + \quad _{SMB}\, SMB_t + \quad _{HML}\, HML_t + \quad _{PMO}\, PMO_t + \quad _{MS}\, MS_t;$$

where R_t is the anomaly return of one momentum spillover factor to be explained, and the MS (momentum spillover) factor on the right-hand side is another momentum spillover factor used as an additional explanatory variable. We use "CH-4+a speci c MS factor" to indicate the set of explanatory variables used to explain the target momentum spillover factor R_t. For example, CH-4+Sentence3 means the set of explanatory variables includes CH-4 factors plus a news co-mention momentum factor of the Sentence3 type. If the news co-mention momentum factor does explain the momentum spillover factor R_t, then we should observe the alpha of the model becoming smaller and potentially insigni cant after controlling for the news co-mention momentum factor. Additionally, the $_{MS}$ should be signi cantly positive.

Panel A of Table 4 shows the alpha of regressing the time series of the target momentum spillover factor R_t on the CH-4 factors plus a speci c MS factor. The column name corresponds to the name of the target factor R_t, and the row name indicates the set of variables in the augmented CH-4 model. All news co-mention momentum factors are constructed using the same_sentence identi cation strategy, which has shown superior performance. We also present results obtained using the same_article method in Table 20 in the robusteness part, which yields similar outcomes. We nd that most of the news co-mention momentum spillover factors, particularly those using a short identi cation window to establish links, are capable of explaining other momentum spillover factors. For instance, after controlling for the Sentence3 factor, the alphas of all other momentum spillover factors become small and insigni cant. While the CH-4 alpha of the statistical momentum factor is as large as 1.34% (t-statistic=3.35), its CH-4+Sentence_3 alpha becomes only 0.17% (t-statistic=0.43). In the case of the analyst, industry, technology, and concept momentum, alphas even turn negative after the addition of the Sentence3 factor.

The results are similar for the Sentence6, Sentence9, and Sentence12 factors, which all have strong explanatory power for other factors. One special case is the geographic momentum factor. Although it has insignificant alpha in the CH-4+Sentence3 model, the factor still has positive and significant alpha if one controls any of the three longer-term news co-mention factors. This is in line with our previous findings that the strength of news co-mention momentum decreases as the length of the identification window increases. Actually, the Sentence3 momentum factor can explain the news co-mention momentum factor with longer identification windows, but not vice versa.

Panel B of Table 4 reports the factor loadings of other momentum spillover factors on the CH-4+Sentence3 model. All the loading coefficients of the Sentence3 factor are significantly positive, and much bigger than the loadings on CH-4, indicating that all other cross-asset momentum anomalies are explained by their loadings on the Sentence3 news co-mention momentum factor.

As a comparison, we also attempt to explain the news co-mention momentum spillover factors with other cross-momentum factors including the factors based on analyst co-coverage momentum, industry momentum, geographic momentum, technology momentum, statistical momentum, and concept momentum. We find that none of the four news co-mention momentum spillover factors can be explained by those other momentum spillover factors as the alphas all remain large and significant. For example, the momentum spillover via statistical linkages yields a high monthly long-short return and CH-4 alpha, which is the second-highest among all trading strategies. However, the Sentence3 momentum factor still generates an alpha of 1.16% (t-statistic=4.17) per month under the CH-4+Statistical model. Moreover, even the weakest news co-mention momentum spillover factor (i.e., the Sentence12 factor) yields a CH-4+Statistical alpha of 0.51% (t-statistic=2.57) per month. The analyst co-coverage momentum factor, which was found to explain all other cross-momentum factors in the US market, cannot explain either the news co-mention momentum spillover factors or the statistical momentum factor. In the last row of panel A, we include all the six momentum spillover factors that are not related to news as additional explanatory variables. We find that the CH-4+Non_news model is unable to explain the news co-mention momentum spillover factors. This suggests that even when we include all other non-news related momentum spillover factors as additional explanatory variables, they cannot explain the abnormal returns associated with the news co-mention momentum anomaly.[17]

We conduct several robustness checks, and the detailed results can be found in subsubsection 6.3.3. Firstly, we run regressions with the Sentence3 factor as the sole explanatory variable (CH-4 factors are excluded), which is shown in Table 17. Besides, we construct industry momentum factors based on other industry classifications, and the results are shown in Table 18. Furthermore, we perform spanning tests for different sub-sample periods, including the periods 2006-2020 and 2012-2020. The results are summarized in Table 19. Finally, in Table 20, we convert the co-mention momentum from same_sentence type to same_article type under the four identification windows. Overall, our main findings remain robust to these changes that we made.

[17] In the CH-4+Non_news model, we include all six non-news factors as explanatory variables. However, we recognize the concern that the small sample size of concept momentum could potentially impact our results. To address this concern, we perform a robustness check by excluding the concept momentum variable from the model. Our results are robust to the exclusion of the concept momentum variable.

Table 4: Factor spanning test

Panel A: Alphas of factor spanning tests

	Sentence_3	Sentence_6	Sentence_9	Sentence_12	Analyst	Industry	Geographic	Technology	Statistical	Concept
CH-4+Sentence _3		-0.15	-0.22	-0.30	-0.44	-0.53	0.25	-0.26	0.17	-0.51
		(-1.51)	(-1.76)	(-2.15)	(-1.76)	(-2.10)	(1.92)	(-1.11)	(0.43)	(-1.59)
CH-4+Sentence _6	0.37		-0.08	-0.16	-0.34	-0.45	0.27	-0.21	0.26	-0.41
	(3.48)		(-1.17)	(-1.80)	(-1.47)	(-1.93)	(2.11)	(-0.95)	(0.71)	(-1.37)
CH-4+Sentence _9	0.54	0.16		-0.08	-0.25	-0.38	0.26	-0.17	0.30	-0.31
	(3.95)	(2.16)		(-1.62)	(-1.13)	(-1.76)	(2.08)	(-0.79)	(0.91)	(-1.10)
CH-4+Sentence _12	0.65	0.27	0.11		-0.17	-0.32	0.27	-0.13	0.36	-0.26
	(4.27)	(2.88)	(2.30)		(-0.81)	(-1.53)	(2.15)	(-0.60)	(1.13)	(-0.94)
CH-4+Analyst	1.18	0.80	0.67	0.58		0.11	0.24	0.17	0.70	0.06
	(4.97)	(4.25)	(3.79)	(3.47)		(0.76)	(1.83)	(0.99)	(2.18)	(0.28)
CH-4+Industry	1.35	0.98	0.84	0.76	0.15		0.25	0.14	0.78	0.33
	(5.08)	(4.49)	(4.30)	(4.03)	(0.96)		(1.92)	(0.87)	(2.25)	(1.67)
CH-4+Geographic	1.62	1.28	1.13	1.06	0.59	0.55		0.44	1.05	0.34
	(4.94)	(4.38)	(3.94)	(3.72)	(2.21)	(2.26)		(2.07)	(2.61)	(0.92)
CH-4+Technology	1.49	1.12	0.99	0.91	0.34	0.27	0.27		0.93	0.25
	(4.63)	(4.07)	(3.86)	(3.63)	(1.59)	(1.41)	(2.05)		(2.50)	(0.98)
CH-4+Statistical	1.16	0.77	0.61	0.51	0.20	0.23	0.23	0.25		0.04
	(4.17)	(3.28)	(2.90)	(2.57)	(0.90)	(0.97)	(1.69)	(1.24)		(0.14)
CH-4+Concept	1.33	0.89	0.71	0.60	0.29	-0.16	0.35	0.04	0.92	
	(4.06)	(3.65)	(3.08)	(2.81)	(1.30)	(-0.89)	(2.19)	(0.18)	(2.61)	
CH-4+Non _news	1.15	0.72	0.52	0.40						
	(3.59)	(3.00)	(2.22)	(1.92)						

Panel B: CH-4+Sentence _3 factor loading

	Sentence_3	Sentence_6	Sentence_9	Sentence_12	Anayst	Industry	Geographic	Technology	Statistical	Concept
Sentence_3		0.896	0.878	0.886	0.835	0.748	0.086	0.518	0.862	0.747
		(28.82)	(23.66)	(21.14)	(11.93)	(10.92)	(4.09)	(7.12)	(8.50)	(6.68)
mktrf		-0.012	-0.014	-0.028	-0.085	-0.069	0.011	-0.034	-0.078	-0.012
		(-1.10)	(-0.82)	(-1.54)	(-3.55)	(-2.60)	(0.84)	(-1.42)	(-1.73)	(-0.30)
VMG		-0.008	-0.008	-0.007	-0.001	-0.018	-0.027	-0.051	0.019	0.004
		(-0.36)	(-0.25)	(-0.21)	(-0.02)	(-0.26)	(-1.00)	(-0.73)	(0.17)	(0.04)
SMB		0.010	0.006	0.016	0.044	0.054	-0.001	-0.026	0.068	0.025
		(0.46)	(0.19)	(0.42)	(0.75)	(0.85)	(-0.05)	(-0.46)	(0.61)	(0.26)
PMO		-0.032	-0.037	-0.057	-0.082	-0.036	-0.018	-0.059	-0.203	0.087
		(-2.03)	(-1.97)	(-2.05)	(-2.06)	(-0.79)	(-1.16)	(-1.49)	(-2.33)	(0.97)
Adj. R2		0.881	0.824	0.797	0.525	0.460	0.063	0.275	0.355	0.351
# Obs.		464	464	464	464	464	464	464	464	227

Panel A reports the intercept (or alpha) of regressing the time series of di erent long-short momentum returns on the CH-4 factors (MKT; SMB; HML; PMO) (Liu et al., 2019) plus each MS cross- rm momentum factor. The construction of the MS factor is given in subsection 4.1. The column name indicates the type of long-short momentum return (i.e., the dependent variable), while the row name means the augmented CH-4 model (i.e., the original CH-4 model plus a speci c MS factor). For example, the alphas of the row CH-4+Sentence _3 indicate the alphas from the time-series regressions of long-short portfolio returns of other momentums on MKT; SMB; HML; PMO and the Sentence_3 factor. In the last row of panel A, we include all the six momentum spillover factors that are not related to news as additional explanatory variables. All news co-mention momentum factors are based on same_sentence type under identi cation windows including 3-month, 6-month, 9-month, and 12-month. The industry momentum is based on the Shenwan-1 classi cation, while the geographic momentum is based on the province level. We do not consider the customer momentum due to the low data quality. The sample period of news co-mention momentum is 2012-2020. The sample period of concept momentum is Aug. 2016-2020. The sample periods for other momentums are 2006-2020. Newey and West (1987) adjusted t-statistics are shown in parentheses. Alphas with t-statistics higher than 2.00 are highlighted in bold. Panel B reports the factor loadings of other momentum spillover long-short returns on the CH-4+Sentence _3 model. Coe cients with t-statistics higher than 2.00 are highlighted in bold.

4.2 Fama-Macbeth Regression Tests

Our spanning tests reveal that all other cross-asset momentum anomalies can be explained by the news co-mention momentum spillover factor. To provide additional evidence, we employ Fama-MacBeth regressions to examine the relationship between the stock excess return in the next week and the weighted average peer rm returns constructed using di erent linkage proxies.[18] Our control variables include the size (taking logarithms), book-to-market ratio, and stock return in the past week. All independent variables are standardized by their cross-sectional means and standard deviations.

We estimate two sets of regressions based on two di erent samples. The rst one takes the union of all samples of average peer rm returns constructed using di erent linkage proxies,[19] and the second sample takes the intersection set of all peer return samples. Table 5 presents the results of the two sets of Fama-MacBeth regressions. Panel A and Panel B present the results for the union sample and intersection sample, respectively. Columns 1-7 use each peer rm return alone as the dependent variable respectively. The past-one-week news co-mention peer rm return, Sentence_3_Rtn, exhibits the strongest predictability power for future returns, both statistically and economically. A one standard deviation increase in the past-one-week Sentence_3_Rtn is associated with an increase of 27.1 bps in the future return, with a t-statistic of 5.53. For the analyst and statistical momentum, which follow the news co-mention momentum according to portfolio analysis, their predictive powers are much smaller, with just 7.1 bps (t-statistic=4.32) and 3.9 bps (t-statistic=2.57) per standard deviation change of the peer rm return, respectively. The geographic peer rm return does not exhibit a signi cant predictive power of the future return based on either the union sample or the intersection